TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY
Julkaisu 737 • Publication 737

Mehdi Rezaei

Advances on Video Coding Algorithms for Streaming Applications

Tampere 2008

Tampereen teknillinen yliopisto. Julkaisu 737
Tampere University of Technology. Publication 737

Mehdi Rezaei

Advances on Video Coding Algorithms for Streaming Applications

Thesis for the degree of Doctor of Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB109, at Tampere University of Technology, on the 24th of May 2008, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2008

ISBN 978-952-15-1981-9 (printed)
ISBN 978-952-15-1993-2 (PDF)
ISSN 1459-2045

Tampereen Yliopistopaino Oy, 2008

Abstract

Video streaming refers to a method for delivery of video services in which compressed video data are transmitted as a continuous stream to the receiver and played as they arrive. Therefore, a transmission channel with enough bandwidth for the continuous transmission of the bit stream should be assured. In video streaming, the quality of service depends on the video encoding process and the allocated bandwidth to the bit stream. The encoding process is very important, because it has a great impact on the rate-distortion performance and also on the utilization of different resources such as processing power, transmission bandwidth, and end-to-end delay of streaming service. In real-time video streaming, compromising the rate-distortion performance, computational complexity, bandwidth efficiency, and the quality of service makes the encoding process a challenge. The challenge can be more serious when the overall performance over a number of streaming services is considered.

The specifications of most video coding standards define only the bit-stream syntax and the decoding process. The encoding process is not standardized to allow flexible implementations. This thesis proposes algorithms for the encoding process with emphasis on rate control for streaming applications.

In this thesis, novel fuzzy logic controllers have been developed and used in several video encoding and streaming scenarios. Video rate control algorithms for encoding variable bit rate video are proposed where the fuzzy controllers are deployed. The proposed algorithms in the new approach are very different from the conventional rate control algorithms in the way they operate. However, some theoretical and practical results of conventional rate control are used in the new approach. Furthermore, video encoding techniques optimized for streaming over DVB-H (Digital Video Coding for Handheld) are proposed in this thesis.

The contributions of thesis are presented in three parts. Each part includes summary of a number of published papers and a selected subset of published papers is attached to the thesis. In the first part of the thesis, video rate control algorithms and tools for video streaming applications are proposed. The proposed rate control algorithms can be applied to other variable bit rate video applications such as recording on storage media. In the second part, video encoding and rate control algorithms are proposed and optimized for video streaming over DVB-H channels. The third part presents video encoding algorithms for a

digital video broadcast system in which a number of video sources are encoded and broadcasted simultaneously.

The proposed rate control algorithms in the thesis outperform sophisticate conventional rate control algorithms in term of rate-distortion performance and computational complexities. Moreover, they maximize utilization of the allocated resource to increase the overall quality of service. Furthermore, the proposed algorithms are very flexible so that they can be easily tuned for different operating points and applications.

Acknowledgments

The research presented in this thesis has been carried out at the Department of Signal Processing, Tampere University of Technology, Finland, in collaboration with the Nokia Research Center, Tampere, during 2003-2008.

First and foremost I wish to express my deepest gratitude to my supervisor, Prof. Moncef Gabbouj, for his excellent guidance and unfaltering support. It was a great privilege and honor to study under his guidance. His unconditional assistance was most crucial to make this dissertation.

I would like to thank the thesis reviewers Prof. Olli Silven and Prof. Fernando Pereira for their review and constructive comments on the manuscript of the thesis.

I would like to thank all my colleagues at Tampere University of Technology and Nokia Research Center for sharing with me their ideas and knowledge. I gratefully thank Miska Hannuksela, Dr. Imed Bouazizi, Prof. Stephan Wenger and Vinod Kumar Malamal Vadakital for their contribution to a pleasant research environment.

I would like to thank the professors of the Department of Signal Processing, especially Prof. Jaakko Astola, Prof. Ari Visa, Prof. Karen Egiazarian, Prof. Irek Defee, Prof. Ioan Tabus, Prof. Tapio Saramäki and Prof. Ulla Ruotsalainen for creating such an open and enjoyable research environment.

I am thankful to all my teachers and professors who educated me from pre-school to university, especially Prof. Mojtaba Lotfizad and Prof. Ehsanollah Kabir. They create enthusiasm in me toward the scientific field of Signal Processing.

I greatly appreciate the help of the staff of Tampere University of Technology, as they made life and work so much easier. I thank all the system administrators Maarit Ahonen-Lehmus, Pentti Jokela, Juha Peltonen, Tommi Kuisma, Sami Lanu, Jari Pärssinen, and Antti Orava, for providing the secure and high performance computer network and equipments. I very warmly thank the outstanding secretaries: Virve Larmila, Ulla Siltaloppi, Elina Orava, Kirsi Järnström, and Pirkko Ruotsalainen for their cordiality and prompt assistance with all administrative matters.

I'm thankful to all my friends for the gift of their friendship. In particular, I was blessed to have friends like Dr. Peyman Arian and Dr. Sassan Iraji in Finland.

During the past years I had the pleasure of sharing my room with Dr. Dong Tian, Ying Chen and Dr. Junyong You. I would like to thank all of them for providing an enjoyable and relaxed workroom atmosphere.

The financial support of Iranian Ministry of Science, Research & Technology and University of Sistan & Balouchestan is gratefully acknowledged. I also thankfully acknowledge Nokia Foundation Award in 2005 and 2006, and TTY Tuula & Yrjö Neuvo's Foundation Award in 2007.

Finally, but by no means the least, I wish to express my sincere gratitude to my big family, especially my wife Farzaneh Fereidoonfar and my son Sina for their incessant love and supports. I owe my gratitude to my sisters Bigom, Tahereh, Masoumeh, and Zahra and their families and to my brothers Ali, Mohammad, Reza, Hasan, and Sadegh and their families. I am also indebted to my parents-in-law Nesa and Hasan, sisters-in-law Azadeh, Behnaz, and Behjat and brother-in-law Hamid for their patience, love and support. I dedicate this thesis to the memory of my parents Fatemeh Yazdanpanah and Hosain Rezaei.

Tampere, May 2008

Mehdi Rezaei

Contents

List of Publications

The thesis is composed of a summary part and 9 publications listed below which are included as appendices.

[P1] M. Rezaei, S. Wenger, M. Gabbouj, “Video Rate Control for Streaming and Local Recording Optimized for Mobile Devices,” *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC),* Berlin, Germany, Sep. 2005, Vol. 4, pp. 2284-2288.

[P2] M. Rezaei, S. Wenger, M. Gabbouj, “Analyzed Rate Distortion Model in Standard Video Codecs for Rate Control,” *IEEE Workshop on Signal Processing Systems (SIPS)*, Athens, Greece, Nov. 2005, pp. 550-555 .

[P3] M. Rezaei, M. Gabbouj, “Bit Allocation for Variable Bitrate Video,” *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Beijing, China, Oct. 2005, pp. 1-4.

[P4] M. Rezaei, M. M. Hannuksela, M. Gabbouj, “Region Based Minimization of Quantization Error Propagation in Variable Bitrate Video,” *Picture Coding Symposium (PCS),* Beijing, China, Apr. 2006, pp. 1-5.

[P5] M. Rezaei, M. M. Hannuksela, M. Gabbouj, “Semi-Fuzzy Rate Controller for Variable Bit Rate Video,” *IEEE Transactions on Circuits and Systems for Video Technology*”, Vol. 18, No. 5, pp. 633-645, May 2008.

[P6] M. Rezaei, I. Bouazizi, V. K. M. Vadakital, M. Gabbouj, “Optimal Channel Changing Delay for Mobile TV over DVB-H,” *IEEE International Conference on Portable Information Devices (Portable)*, Orlando, USA, Mar. 2007, pp. 1-5.

[P7] M. Rezaei, M. M. Hannuksela, M. Gabbouj, “Tune-in Time Reduction in Video Streaming over DVB-H,” *IEEE Transactions on Broadcasting, Special Issue on "Mobile Multimedia Broadcasting",* Vol. 53, No. 1, Part 2, pp. 320-328, March 2007.

[P8] M. Rezaei, I. Bouazizi, M. Gabbouj, “Joint Video Coding and Statistical Multiplexing for Broadcasting over DVB-H,” *IEEE Transactions on Multimedia*”, In Press, 2008.

[P9] M. Rezaei, I. Bouazizi, M. Gabbouj, “Fuzzy Joint Encoding and Statistical Multiplexing of Multiple Video Sources with Independent Quality of Services for Streaming over DVB-H,” *IEEE Int. Con. Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007)*, Kaohsiung, Taiwan, Vol. 1, Nov. 2007, pp. 542-545.

List of Supplementary Publications

The contents of this thesis are also closely related to the following publications by the author:

[S1] M. Rezaei, A. Akhbardeh, M. M. Hannuksela, M. Gabbouj, "Fuzzy Rate Controller for Variable Bitrate Video in Mobile Applications," *IEEE International Conference on Communications (ICC),* Istanbul, Turkey, June 2006, Vol. 7, pp. 3197-3201.

[S2] M. Rezaei, M. M. Hannuksela, M. Gabbouj, "Low-Complexity Fuzzy Video Rate Controller for Streaming," *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Toulouse, France, May 2006, Vol. 2, pp. 897-900.

[S3] M. Rezaei, M. Gabbouj, I. Bouazizi, "Delay Constrained Fuzzy Rate Control for Video Streaming over DVB-H," *IEEE Int. Con. Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP),* Pasadena, California, USA, Dec. 2006, pp. 223-227.

[S4] I. Bouazizi, M. Rezaei, "Approaches for Smooth Buffering in Mobile TV over DVB-H," *56th Annual IEEE Broadcast Symposium*, Washington, USA, Sep. 2006, pp. 1-4.

[S5] M. M. Hannuksela, M. Rezaei, M. Gabbouj, "Video Splicing for Tune-in Time Reduction in IP Datacasting over DVB-H," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (ISBMSB),* Las Vegas, USA, Apr. 2006, pp. 1-6.

[S6] M. Rezaei, M. M. Hannuksela, M. Gabbouj, "Video Encoding and Splicing for Tune-in Time Reduction in IP Datacasting (IPDC) over DVB-H," *IEEE International Conference on Multimedia & Expo (ICME)*, Toronto, Canada, July 2006, pp. 601-604.

[S7] M. Rezaei, M. M. Hannuksela, M. Gabbouj, "Video splicing and Fuzzy Rate Control in IP Multi-Protocol Encapsulator for Tune-in Time Reduction in IP Datacasting (IPDC) over DVB-H," *IEEE International Conference on Image Processing (ICIP)*, Atlanta, USA, Oct. 2006, pp. 3041-3044.

[S8] M. Rezaei, M. M. Hannuksela, V. K. M. Vadakital, M. Gabbouj, "Spliced Video and Buffering Considerations for Tune-in Time Minimization in DVB-H for Mobile TV,"

IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Helsinki, Finland, Sep. 2006, pp. 1-5.

List of Abbreviations

AVC	Advanced Video Coding standard
CBR	Constant Bit Rate
CERCS	Content Encoder Rate Control System
DCT	Discrete Cosine Transform
DRBS	Decoder Refresh Bit Stream
DVB	Digital Video Broadcast
DVB-H	Digital Video Broadcast for Handheld
ETSI	European Telecommunications Standards Institute
FEC	Forward Error Correction
FGS	Fine-Granular Scalable
GOP	Group of Pictures
HRD	Hypothetical Reference Decoder
IDR	Instantaneous Decoder Refresh
IERCS	IP Encapsulating Rate Control System
IP	Internet Protocol
IPDC	Internet Protocol Data Cast
IRC	Independent Rate Controller
JRC	Joint Rate Controller
JVRC	Joint Video Rate Controller
JVT	Joint Video Team
MAD	Mean Absolute Difference
MB	Macroblock
MMS	Multimedia Messaging Services
MPE	Multiprotocol Encapsulation
MPEG	Moving Picture Experts Group
MSE	Mean Square Error
MSF	Membership Function
PSNR	Peak Signal to Noise Ratio
QoS	Quality of Service
QP	Quantization Parameter
RCA	Rate Control Algorithm
RD	Rate-Distortion
ROI	Region of Interest

RTP	Real Time Protocol
SAD	Sum of Absolute Difference
SBS	Spliceable Bit Stream
SEI	Supplemental Enhancement Information
SI	Switching I picture
SOF	Set of Frames
SP	Switching P picture
SPP	Special P-frame
SPPMB	Special P Macroblock
SRC	Scalable Rate Control
StatMux	Statistical Multiplexing
SVC	Scalable Video Coding
TBRE	Target Bit Rate Estimator
TMN	Test Model Near-term
VBR	Variable Bit Rate

List of Symbols

O_B	Buffer Occupancy or Fullness
B_S, S_B	Buffer Size
TR, R_T	Target Bit Rate
R	Rate
Q	Quantization Parameter
$H(z)$	Filter Impulse Response
D	Distortion
FR, F	Frame Rate
x_1, x_2	Fuzzy Inputs
$f(x_1, x_2)$	Output of Fuzzy System
$A_j^{i_j}$	Fuzzy Set
$\mu_{A_j^{i_j}}$	Fuzzy Membership Function
$\overline{y}^{i_1 i_2}$	Desired Central Values of Fuzzy Output

Chapter 1

Video Coding and Rate Control

Video coding refers to reducing the quantity of data used to represent a sequence of video pictures or frames. A number of standards for video coding have been defined and research for new video coding standards is ongoing. MPEG-2 [ISO/IEC 1994], H.263 [ITU-T 1995], MPEG-4 [ISO/IEC 1999] and H.264/AVC (Advance Video Coding) [JVT 2003], [Wiegand 2003b] are samples of widely used video coding standards. Many of video coding standards including all ISO/IEC JTC1 and ITU-T standards since H.261 [ITU-T 1990] use a common approach so-called *block-based hybrid* video coding. The basic coding algorithm is a hybrid of inter-picture prediction using temporal redundancy and transform coding of the prediction error signal to reduce spatial redundancy. Each video picture partitioned into fixed-size macroblocks (MB) of 16×16 pixels, which can be encoded in several modes. All hybrid video coding standards define *intra coded* pictures and *predictive coded* pictures. In an Intra coded picture, all MBs are coded without referring to other pictures in the video sequence. In predictive coded pictures, the MBs are typically encoded by a variety of inter coding modes.

Although MPEG-2, H.263, MPEG-4 (part 2), and H.264/AVC video coding standards define similar coding tools, they differ in some enhancement features [Sullivan 2005], [Richardson 2005]. The main differences include the construction of the prediction signal, the block sizes used for transform coding, and the entropy coding methods.

The specifications of most video coding standards provide only the bit-stream syntax and the decoding process in order to enable interoperability. The encoding process is not standardized to permit flexible implementations suitable for different applications. To manage the large number of coding tools included in standards and the broad range of applications, the concept of profiles and levels is employed. A profile defines a set of coding tools that can be used for generating a standard compliant bit stream and a level imposes constraints on certain key parameters of the bit stream, such as the picture resolution, bit rate and MB processing rate.

The standard specifications do not define the encoding algorithm exactly. A video coding algorithm defines how to use the standard tools to encode a video source for a specific application. The choice of video coding algorithm and encoding parameters affect the coded bit rate and the quality of decoded video sequence as well as the computational complexity of the video codec. The precise relationship between coding parameters, bit rate, and visual quality varies depending on the characteristics of the video content such as motion and texture properties. At the same time, practical limits put some constraints on the bit rate, picture quality, and computational complexity that may be achieved.

It is important to control the video encoding process in order to maximize compression performance while the practical constraints are achieved. Rate, distortion, delay, processing complexity, scalability, and error resiliency are typical practical constraints. For example, a multi-pass offline high complexity encoding process with quality constraint can be used for video recording applications while for wireless conversational applications a real-time, low delay, low complexity, error resilient encoding with rate constraint is needed.

1.1. VIDEO RATE CONTROL

The bit rate of an encoded video is considered as a constraint especially in video communication applications. The bit rate of an encoded bit stream that is defined by the video coding algorithm can be controlled by a rate control algorithm (RCA). The RCA attempts to maximize the quality of a compressed video subject to the bit rate and other practical constraints. While most of encoding parameters defined by the video coding algorithm affect the rate-distortion performance, the video rate control algorithm can be assumed as a major part of the video coding algorithm.

A great deal of attention has been paid to video rate control over the past two decades. Many alternative RCAs exist; sophisticated algorithms can achieve excellent rate-distortion performance, usually at a cost of increased computational complexity. The careful selection and implementation of a RCA can make a big difference on encoding performance.

Various rate control algorithms perform the rate control on different levels such as scene, GOP (group of pictures), frame, slice, and MB level. In each level only some encoding parameters are adjustable. The encoding parameter that is most widely used for rate control purposes is the quantization parameter (QP) which could be changed in different levels. As a usual approach, control algorithms operate in two steps. In the first step, a bit budget is allocated to a video segment such as GOP, frame, or a MB according to practical constraints and video properties. In the second step, a QP is computed according to the allocated bit budget and the coding complexity of video. Usually a rate-distortion (RD) model is utilized for computation of QP. The RD model is derived analytically or empirically. In analytical modeling, a RD model is derived according to the statistics of the source video signal and the properties of the encoder. Empirical modeling attempts to approximate the RD curve by interpolating between a set of sample points. The RD model provided by one of the two approaches is then employed in calculation of the QP for the rate control. The RD model parameters are updated according to encoding results. A number of used RD model and rate control algorithms are reviewed in the following sections.

1.2. FROM INFORMATION THEORY TO OPERATIONAL RD MODELS

Rate distortion theory was created by Claude Shannon in his foundational work on information theory [Shannon 1948]. Rate distortion theory determines the minimal amount of information R that should be communicated over a channel, so that the source (input signal) can be approximately reconstructed at the receiver (output signal) without exceeding a given distortion D. Rate distortion theory gives theoretical bounds for the amount of compression that can be achieved by lossy data compression methods. Many of the existing image and video compression techniques that have transforms, quantization, and bit rate allocation procedures can exploit the general shape of RD models.

A RD model is one of the key elements in video encoding and rate control task. Optimizing rate, distortion and other resources require a deep insight into RD function. Video rate is mainly controlled by the QP which is related to the distortion in a decoded video. An accurate description of the RD model based on QP makes it possible to control the bit rate of a compressed video precisely. In this section a number of deployed RD models are reviewed. We try to show how the many available RD models are similarly related to basic theories and also to each other.

In hybrid video codecs the information is quantized in a transform domain such as DCT (discrete cosine transform) domain. Transforming video data to a reversible transform domain does not change the amount of information and distortion in the data [Andrews 1968]. Gish [Gish 1968] proved that for a random variable with a reasonably smooth density function that is quantized by a uniform quantizer the minimum entropy of the quantized signal is:

$$H_{\min} \approx H_0 - \log_2 \delta \tag{1.1}$$

where H_0 denotes the entropy of a continuous distribution and δ is the quantization step. Also the distortion *D* for any error function of $L(x-\hat{x})$ is:

$$D \approx M(\delta) \tag{1.2}$$

where

$$M(\delta) = \frac{1}{\delta} \int_{-\delta/2}^{\delta/2} L(u)du \tag{1.3}$$

Combining the results above yields:

$$H_{\min} \approx H_0 - \log_2(M^{-1}(D)) \tag{1.4}$$

which for a Gaussian distribution and *Mean Square Error* (MSE) then leads to:

$$D = \delta^2 / 12 \tag{1.5}$$

$$H_{\min} \approx H_0 - (1/2)\log_2(12D) \tag{1.6}$$

or it can be written as

$$H_{\min} \approx (1/2)\log_2(V_0 / D) + (1/2)\log_2(2\pi e / 12) \tag{1.7}$$

or as

$$H_{\min} \approx (1/2)\log_2(V_0 / D) + 0.255 \tag{1.8}$$

where e denotes the natural number and V_0 is the variance of a Gaussian distribution that has the same entropy H_0. This means that the uniform quantizer can achieve a performance asymptotically within approximately ¼ bit of the Shannon RD lower bound [Shannon 1948]. Similar results can be found in [Jayant 1984], [Gersho 1992]. According to these results for uniform quantizer, at high bit rate (low distortion), the rate and distortion models can be summarized as [Hang 1997]

$$R(\delta) = \frac{1}{2}\log_2\left(\frac{\varepsilon^2 \beta \sigma_x^{\ 2}}{\delta^2}\right) \tag{1.9}$$

$$D(\delta) = \frac{\delta^2}{\beta} \tag{1.10}$$

where β is 12 for small δ. ε^2 is a source dependent coefficient that is about one for uniform, 1.4 for Gaussian, and 1.2 for Laplacian distributions. $\sigma_x^{\ 2}$ is the signal variance. To account for a wide range of δ, parameter β typically needs to be empirically adjusted based on samples of RD curve. For low bit rate where a higher compression ratio is obtained by a coarser quantization and larger quantization steps more accurate RD models have been proposed in the literatures.

In the case of a uniform quantization of the Laplacian distribution

$$p(x) = \frac{1}{\sqrt{2}\sigma}\exp\left(-\frac{\sqrt{2}\,|x|}{\sigma}\right), \tag{1.11}$$

with a quantization step size Q, the entropy of the quantized signal can be derived as [Moscheni 1993]

$$H = -p(0)\log_2(p(0)) - 2\sinh\left(\frac{\alpha}{2}\right)\left[\log_2\left(\sinh\left(\frac{\alpha}{2}\right)\right)\frac{1}{e^{\alpha}-1} + \frac{\alpha}{\ln 2}\frac{-e^{\alpha}}{\left(e^{\alpha}-1\right)^2}\right], \quad (1.12)$$

where

$$\alpha = \sqrt{2}Q/\sigma \ . \quad (1.13)$$

A Laplacian distribution for DCT coefficients was reported by Smoot [Smoot 1996]. An approximation of this solution has been used as [Ribas-Corbera 1996-9]

$$H(Q) = \begin{cases} \frac{1}{2}\log_2\left(2e^2\frac{\sigma^2}{Q^2}\right) & ,\frac{\sigma^2}{Q^2} > \frac{1}{2e} \\ \frac{e}{\ln 2}\frac{\sigma^2}{Q^2} & ,\frac{\sigma^2}{Q^2} \le \frac{1}{2e} \end{cases}, \quad (1.14)$$

where σ^2/Q^2 is larger than $1/2e$ for the low distortion (high rate) case and smaller than $1/2e$ for the low rate.

For Laplacian sources with density $p(x) = (\lambda/2)e^{-\lambda|x|}$, the RD function can also be written in terms of the *Mean Absolute Difference* (MAD) distortion D [Berger 1971]:

$$R(D) = -\ln\left(\lambda D\right), \quad (1.15)$$

where $D_{\min} = 0$, $D_{\max} = 1/\lambda$, $0 < D < 1/\lambda$ and λ is a constant. Chiang *et al.* [Chiang 1997] used Taylor expansion of this model to derive an operational RD model as:

$$R(D) = \left(\frac{1}{\lambda D} - 1\right) - \frac{1}{2}\left(\frac{1}{\lambda D} - 1\right)^2 + R_3(D), \quad (1.16)$$

$$R(D) = -\frac{3}{2} + \frac{2}{\lambda}D^{-1} - \frac{1}{2\lambda^2}D^{-2} + R_3(D) \ . \quad (1.17)$$

Accordingly, they proposed a rate model with respect to quantization parameter Q and applied it to the MSE distortion as [Chiang 1997]

$$R(Q) = aQ^{-1} + bQ^{-2} \ , \quad (1.18)$$

where a and b are model parameters which are updated during encoding. Using MAD distortion and considering overhead information, this model has been enhanced as [Lee 2000]

$$R = M\left(aQ^{-1} + bQ^{-2}\right) + H, \quad (1.19)$$

where M denotes MAD and H stands for overhead information.

Kamaci *et al*. [Kamaci 2005] proposed a zero-mean Cauchy density function for the source with parameter μ as

$$p(x) = \frac{1}{\pi}\frac{\mu}{\mu^2 + x^2}, \tag{1.20}$$

and a RD model as

$$D = cR^{-\gamma}, \tag{1.21}$$

where c and γ are some properly selected constants.

In another approach, He *et al*. [Kim 2001], [He 2001-3], proposed a unified $\rho-domain$ RD model, in which the bit rate is estimated by a linear function of the percentage of zero coefficients in each video frame or MB. The proposed linear model is

$$R(\rho) = \theta\left(1-\rho\right), \tag{1.22}$$

where ρ is the percentage of zeros among the quantized transform coefficients and is the only parameter of the model which is estimated from the coding statistics.

Chang *et al.* [Chang 2006] proposed a new RD model in $q-domain$ that estimates the encoding bit rates based on the number of nonzero coefficients, the count of zeros before the last nonzero coefficient in the zigzag-scan order, and the sum of absolute quantized nonzero coefficients.

Dai *et al*. [Dai 2003] proposed a mixture of Laplacian distribution for MPEG-4 *Fine-Granular Scalable* (FGS) video as:

$$p(x) = q\frac{\lambda_0}{2}e^{-\lambda_0|x|} + (1-q)\frac{\lambda_1}{2}e^{-\lambda_1|x|}, \tag{1.23}$$

where the random variable x represents the DCT residue, q is the probability to obtain a sample from one of two Laplacian components and λ_0 and λ_1 are the shape parameters of a Laplacian distribution. For bit plane coding with quantization step Δ, the MSE distortion for a Laplacian distribution is computed as:

$$D(\Delta) \approx \frac{2ae^{b(\Delta-1)}}{1-e^{b\Delta}}\left(\frac{\left(\Delta-1\right)^2}{b} - \frac{2\left(\Delta-1\right)^2}{b^2} + \frac{2}{b^3}\right), \tag{1.24}$$

where a and b are constants that depend on the distribution parameters. The distortion function for the mixture of Laplacian will be a linear combination of two functions as above. Dai *et al*. [Dai 2006] combined this distortion model with the $\rho-domain$ RD model and proposed a R-D model based on PSNR (Peak Signal to Noise Ratio) distortion measure as:

$$PSNR(R) = AR + B\sqrt{R} + C, \tag{1.25}$$

where A and B are estimated from at least two (R,D) samples and $C = 10\log_{10}\left(255^2 / \sigma_x^{\ 2}\right)$. For uncorrelated (or weakly correlated) sources $\sigma_x^{\ 2}$ is the variance of the source.

Analytical RD models mainly differ in assumed distribution for signal and the used distortion measure for quantization error. Besides the above RD models, there are purely empirical ways to estimate RD curves. Among the numerous studies, e.g., Lin *et al*. [Lin 1998] used a cubic interpolation of the empirical curve and Zhao *et al*. [Zhao 2002] applied similar methods to FGS video.

1.3. OPERATIONAL RATE CONTROL ALGORITHMS

A number of RD models were reviewed in the previous section. In this section a general approach in which the RD models are used for video rate control is explained and then a number of operational RCAs are reviewed. Several encoding parameters such as picture size, picture type (Inter or Intra), frame rate, and QP affect the RD performance and can be used to control the bit rate. However, the quantization parameter is the main controlling parameter that is used for the rate control. In this section, QP is also considered as the main controlling parameter and it is explained how the quantization parameter is determined by various RCAs.

Figure 1 shows a general block diagram for video rate control. As shown, the uncompressed video source is compressed to an encoded bit stream by the video encoder. Generally, the encoded bit stream is stored in a local media storage or it is transmitted through a channel. The stored bit stream later on can be transmitted to a channel or it can be decoded and played out. When the compressed video is going to be transmitted via a channel in real-time or even after recording locally, usually the bit stream is constrained to some buffering parameters to guarantee a continuous play out with regards to variations in bit rate of bit stream and variations in band-width of the transmission channel.

To impose the buffering constraint on the bit stream, a buffer with desired parameters is used while encoding. The buffer can be a real buffer that is located between the encoder and transmitter in real-time applications or it can be a virtual buffer in offline applications that simulates a real buffer. Buffer information including size and occupancy is used by the rate controller as a feedback signal. In case of real-time applications, some information about the channel can be used by the rate control as feedback. If the compressed video is targeted only for media storage, the buffering constraint may be unnecessary. In this case, the storage size can be considered as a constraint while the occupancy of storage may be used as a feedback signal by the rate controller. A rate controller can also use some information about the rate and distortion of encoded bit stream as feedback.

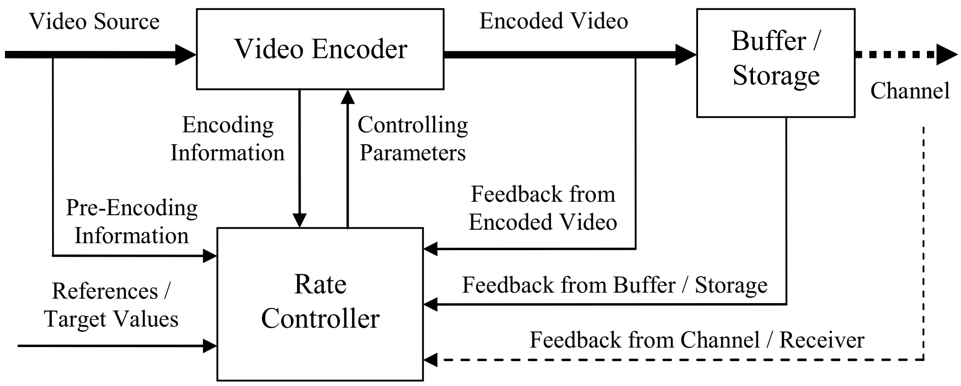

Figure 1: A simple block diagram of a video rate control

Beside feedback information, some feed forward information obtained from uncompressed video can be used by the rate controller. Feed forward information is usually collected by look ahead and preprocessing of uncompressed video. Look ahead can be limited to a short period such as a number of MBs at the same video frame or it can be limited to a longer period such as a GOP or even the whole sequence in multi-pass encoding. Some target values such as average bit rate are defined for the rate controller that are used as references for the rate controller. Finally, some information can be provided by the encoder for the rate controller during encoding.

The rate controller determines the controlling parameters of the encoder based on the received information from different sources and based on a rate control algorithm. Various rate control algorithms differ in their inputs and the relationships that exist between their inputs and outputs. As a widely used approach, the rate controller allocates a bit budget to a video segment (e.g. GOP, frame, and MB) according to the inputs information. Then using a RD model a quantization parameter is computed for encoding the video segment. The parameters of the RD model are updated during the encoding process based on encoding results. Various bit allocation methods and RD models are employed by rate controllers. A number of RCAs are compared in the sequel.

Version 5 of MPEG-2 video *Test Model* describes a RCA for constant bit rate (CBR) encoding that takes into account the different properties of the three coded picture types (I, P and B-pictures) [ISO/IEC 1993]. The algorithm utilizes three virtual buffers corresponding to three types of frames. A simple first order RD model $R = X / Q$ parameterized for each type of frame is used. A bit budget is allocated to a GOP and then to each frame according to the average bit rate and relative complexity of the frame type. The occupancy of the proper virtual buffer is updated after encoding each MB. Then, QP for the next MB is computed and modulated based on buffer occupancy and a special activity measure.

The H.263 *Test Model Version 8* (TMN8) uses a RCA consisting of a control at the frame and MB levels [ITU-T 1997], [Gardos 1997], [Ribas-Corbera 1999]. At the frame level, each encoded frame is added to the encoder buffer and each transmitted frame is removed from the buffer. If the number of bits in the buffer exceeds a threshold *M*, the next frame is skipped; otherwise a target number of bits B is set for encoding the next frame. A higher threshold *M* means fewer skipped frames, but a larger delay through the system. At the MB level, the algorithm operates based on a RD model which is extracted from (1.14) for the low bit rate case as

$$B = A\left(K\frac{\sigma^2}{Q^2} + C\right), \tag{1.26}$$

where Q is the quantization step for the MB (here $Q = 2\,\mathrm{QP}$), A is the number of pixels in a MB (i.e., $\mathrm{A} = 16^2$ pixels), K and C are constants, and σ is the empirical standard deviation of the luminance and chrominance values in the (motion-compensated or intra) MB.

MPEG-4 *verification models* (VM5 or later on) describe an optional rate control algorithm in annex L.1, known as *Scalable Rate Control* (SRC) scheme. This algorithm is appropriate for a single video object and a range of bit rates and spatial/temporal resolutions. The SRC described in Annex L.1 offers rate control at the frame level only i.e. a single QP is chosen for encoding a complete frame. It attempts to achieve a target bit rate over a certain number of frames [Chiang 1995], [Chiang 1997], [ISO/IEC 2001]. The SRC uses the RD model (1.18) as:

$$R = \frac{X_1 S}{Q} + \frac{X_2 S}{Q^2}, \tag{1.27}$$

where Q is the quantization step size (here $Q = 2\mathrm{QP}$), S is the MAD of the residual frame after motion compensation. X_1 and X_2 are the model parameters. S provides a measure of complexity and it is easier to compute than the standard deviation used in H.263 TM8 because the MAD is calculated and used during motion estimation. The SRC consists of the following steps which are carried out after motion compensation and before encoding each frame:

- Calculate a target bit rate R, based on the number of frames in the frame set, the bit budget available for the remainder of frames in the frame set, and the buffer state.
- Calculate S for the complete residual frame.
- Solve the model equation to find the quantization step size Q and QP to be applied to the whole frame.
- Encode the frame.
- Update the model parameters X_1, X_2 based on the actual number of bits generated by the frame.

The SRC algorithm differs from H.263 TM8 in two significant ways: it aims to achieve a target bit rate over a set of frames and it does not change the quantization step within a frame. This can provide a more uniform visual appearance within each frame but makes it difficult to maintain a small buffer size and hence a low delay. An extension to SRC is described in Annex L.3 of MPEG-4 which supports modulation of the quantization step at the MB level and is therefore more suitable for low delay applications. The MB rate control extension is similar to H.263 TM8 rate control.

Joint Video Team (JVT) reference model proposes a RCA for H.264/AVC encoders [Sullivan 2003]. The algorithm is partially based on the proposed algorithms by Ma *et al.* in [Ma 2003a], [Ma 2003b]. The algorithm is used to create the bit stream satisfying the available bandwidth provided by a transmission channel and is also compliant to the standard *hypothetical reference decoder* (HRD) [Ribas-Corbera 2003]. It consists of a tight control in three levels including: GOP level, picture level and an optional basic unit level rate control. The basic unit is defined as a group of successive MB in the same frame. The GOP level rate control calculates the total bits for the uncompressed pictures in the GOP based on picture type, the total number of pictures in the GOP, the instant available bit rate, the occupancy of the virtual buffer, and the actual generated bits by the encoded pictures. The picture level rate control consists of two stages: pre-encoding and post-encoding. The objective of pre-encoding stage is to compute a QP of each picture. The QPs of stored pictures are computed differently from QP of non-stored pictures. The QPs for non-stored pictures are computed by a simple interpolation method on the QPs of stored pictures. The QP for stored picture is computed during three steps. First, a target bit rate for the picture is computed based on buffer state, the frame rate, and the available channel bandwidth. Second, the MAD of the current stored picture is estimated by a linear prediction using the actual MAD of the previous stored picture. Finally, the quantization step corresponding to the target bits is computed using the quadratic RD model:

$$T = C_1 \times \frac{MAD}{Q} + C_2 \times \frac{MAD}{Q^2} - h \quad , \tag{1.28}$$

where h is the total number of header bits and motion vector bits, C_1 and C_2 are two coefficients. The corresponding QP is computed using the relationship between the quantization step and QP of H.264/AVC as $Q = 2^{(QP/6)}$. To maintain smoothness of visual quality among successive frames, the variations of QP is smoothed.

After encoding a picture, parameters of the linear prediction model for MAD estimation, as well as parameters of RD model are updated. A linear regression model is used to update these parameters. Meanwhile, the actual bits generated are added to the buffer. To ensure that the updated buffer occupancy is not too high; a number of pictures may be skipped similar to the method used in MPEG-4. If the basic unit is selected to be smaller than a frame, an additional basic unit layer rate control for the stored picture should be added. The basic unit layer rate control modulates the values of quantization parameters of all basic units in a frame, so that the sum of generated bits is close to the frame target bits.

Beside RCAs used in standard reference models, a large number of RCAs have been proposed for different applications in the literature. However, most of them can be matched to the general block diagram shown in Figure 1. Tsai *et al*. [Tsai 2004] proposed a modification to TMN8 rate control. They investigated the relationship between quantization distortion and the coding order of MBs. Based on the investigation results they modified the encoding order of MB to favour the more complex MBs. Pan *et al*. [Pan 2003] propose some modifications to the MPEG-4 VM5 rate control by a weighted bit allocation to the P frames. A higher bit budget is allocated to the earlier encoded frames which are used as reference by more subsequent frames. Navakitkanok *et al*. [Navakitkanok 2004] improved the performance of JM rate controller in low delay case by decreasing the number of skipped frames.

Ma *et al*. [Ma 2005] proposed another RCA for H.264/AVC revising the RD model used in the JM RCA. Unlike H.263 and the previous standards, the relationship between the quantization step and QP is not linear in H.264 [Malvar 2003]. Considering the nonlinear relationship between QP and quantization step i.e. $Q = 2^{QP/6}$, a first-order model is proposed as:

$$R = K\frac{SAD}{Q} + C\,, \tag{1.29}$$

where R is the estimated number of coded bits of a MB and SAD is the sum of absolute difference of the motion compensated MB. K and C are model parameters that depend on the type of MB (I, P , B). The first term reflects the bits used to code the transform coefficients. The second term represents the bits used to code the header information of a MB. Using the RD model a RCA is presented. The algorithm provides a one-pass control at the frame level and a partial two-pass control at the MB level.

Some RCAs operate based on RD models which are extracted heuristically. He *et al.* [He 2001], [He 2002] proposed RCAs based on the $\rho - domain$ RD model and Chang *et al.* [Chang 2006] proposed a rate control based on the $q - domain$ RD model.

A number of RCAs have been proposed specially for variable bit rate (VBR) applications such as video streaming and local recording applications. The algorithm presented in [Reed 2001] is a low complexity frame-level rate controller for streaming applications. Although this algorithm utilizes a virtual buffer, two other parameters predominantly control its operation: a large time interval and a large bit budget. The virtual buffer, which is essential in streaming, does not play an active role in this algorithm. The SPEM (Smooth Pursuit Eye Movement) rate control scheme introduced in [Nguyen 2002] is designed for real time streaming. This RCA works near the constant bit rate region, and cannot utilize the variable bit rate benefits effectively. The algorithm presented in [Takamura 2002] is a low complexity RCA targeted for recording application. It tries to suppress the fluctuation of QP as much as possible. The buffer constraint, which is essential in streaming applications, is not considered in this algorithm. Two RCAs are proposed in [Jagmohan 2003] and [Song 2003] for storage media applications that tries to satisfy a target bit budget constraint corresponding to the size of the media. Another RCA is proposed in [Kondo 1997] for storage media. It performs bit allocation according to the coding complexity while it does not impose any constraint on the bit rate. Therefore, depending on the content activity, it produces extreme bit rate variations, which do not obey buffering constraints.

A group of RCAs presented in the literature use a kind of look ahead for the rate control, see [Varsa 2001] and [Pao 1998] as examples. These RCAs require more memory for storage of uncompressed video and more time for pre-processing or multi-pass encoding. Therefore, they are not suitable for real-time applications.

The video rate control task can be combined with other concerns in video coding such as perceptual quality [Pickering 1994], [Nguyen 2002], region of interest (ROI) video coding [Song 1999], [Yang 2005], object-based video coding [Ronda 1999], [Vetro 1999] scalable video coding (SVC) [Lee 2000], [Zhang 2003]. Moreover, other video coding parameters such as frame rate [Reed 2001], [Song 2001] and GOP structure [Yoneyama 2001] have been used for the rate control.

In some applications the complexity of RCA can be a serious constraint. Takamura *et al.* 0 [Takamura 2002] proposed a simple one-pass RD model and a low complexity RCA for variable bit rate application. Kim *et al.* [Kim 2003] used a frame level RCA for H.263 to decrease the computational complexity of the algorithm in comparison to the rate control at the MB level. To avoid complex floating-point RD calculation, Tsai [Tsai 2005] used a dynamic table that relates the encoding bit count, encoding complexity, and QP for each MB. The table contains the RD information implicitly and is updated dynamically on an MB basis to reflect the variation of video contents.

1.4. VIDEO CODING AND RATE CONTROL FOR STREAMING APPLICATIONS

In video streaming applications, usually VBR video bit streams are used. The VBR bit streams video generally provide higher visual quality and compression performance than CBR bit streams at the expense of more resources in terms of transmission bandwidth and delay [Lakshman 1998], [Zhang 2003]. In video streaming applications the transmission delay is not very critical as in conversational applications. Moreover, in some video streaming applications such as broadcast applications in which the service is consumed by many users, it is acceptable to allocate a higher bandwidth to compensate the variation in bit rate. Furthermore, in many applications, VBR bit streams can share a transmission channel using statistical multiplexing to improve the utilization of resources. Statistical multiplexing can considerably increase the bandwidth usage and decrease the transmission delay by sharing the resource among the bit streams. Sample results can be found in [Zhang 1997], [Maglaris 1998], [Bashforth 1998].

Encoding VBR video can be open-loop or closed-loop. In open-loop encoding, the video pictures are encoded with an almost constant QP to provide a relative constant quality for encoded video regardless of the coding complexity of video source. In closed-loop encoding, the bit rate of encoded bit stream is controlled by a VBR rate controller according to feedback signals from encoding results and coding complexity of video source. The rate controller imposes some constraints on the degree of variability allowed in the bit rate. The level of variations in the bit rate determines the required resources in terms of bandwidth and delay for a streaming session. Transmission bandwidth and delay are two resources that can compensate each other. For example, the maximum required bandwidth for the transmission of a VBR bit stream decreases by a smoothing buffer while the smoothing buffer imposes a delay on the system. The size of a smoothing buffer can be considered as a constraint for the bit stream that is imposed by the rate controller. A bigger buffer size means that more variations in the bit rate are allowed.

Contributions of this thesis concentrate more on VBR video encoding algorithms for streaming applications. However, some of the proposed algorithms in this thesis can be easily tuned to provide a high performance for a wide range of applications.

1.5. AUTHOR'S CONTRIBUTION TO THE PUBLICATIONS

The thesis and the publications represent original work, of which the author has been the main contributor. In particular, all video rate control algorithms presented in the thesis have been originally proposed and developed by the author. The idea of video splicing introduced in [S5] has been proposed by Miska Hannuksela which has been implemented, developed and combined with video rate control by the author. All the publications, except [S4], have been written by the author and reviewed and edited by co-authors. However, this work would not have been possible without the support and help of expert co-authors as well as of colleagues at the Tampere University of Technology and the Nokia Research Center.

Chapter 2

VBR Video Rate Control for Streaming

A video RCA can operate in different regions in the RD space between the constant rate region and the constant quality region. In this thesis, RCAs are classified into three classes: CBR, VBR, and constant quality RCAs. However, the terms VBR and constant quality are used with different meaning in the literature. An ideal CBR rate control algorithm operates in a region that is parallel to the distortion (or quality) axis and an ideal constant quality RCA operates in a region that is parallel to rate axis. Although in practice some variations in the bit rate of compressed video even in CBR can exist, the term VBR used here means considerable variation in the bit rate. A VBR rate controller operates in a region between the CBR and constant quality operating areas. It can provide more constant quality in comparison with CBR and less variations in bit rate in comparison to the constant quality case. A more constant quality for a bit stream can be achieved by more variations in bit rate that means more transmission and buffering delay.

For a systematic control, it is difficult to find an accurate reference point for the VBR rate controller. In CBR, the target bit rate is a fixed reference point and the main objective of the controller is to drive the bit rate toward the reference point. In constant quality rate control, the quality of the encoded video is a reference point for the controller. In VBR a long-term average value is defined for the bit rate and there is no real short-term reference point in terms of rate or quality for the controller. The objective of a VBR is to minimize the variations in quality while the bit rate can have some variations with a buffer or a delay constraint.

Some RCAs such as MPEG-4 VM5 [ISO/IEC 2001] and H.264/AVC JM [Sullivan 2003] can be tuned to provide a type of VBR bit stream by operating at the frame level. Unlike TMN8 rate controller that is optimized for low delay applications, these algorithms can provide bit streams with high quality I-pictures. However, these algorithms are not optimized for VBR applications because they operate based on only a short-term reference point from the rate. Moreover, it is shown in this thesis that VBR rate control algorithms can have a lower degree of complexity. In this thesis RCAs are proposed that are optimized for VBR applications. A number of contributions on VBR video rate control are summarized in the sequel.

2.1. VIDEO RCA FOR STREAMING AND LOCAL RECORDING

In [P1] a real-time video RCA designed with buffer constraints is proposed that can be used for streaming and also for local recording applications. The proposed algorithm has a low degree of complexity such that it can be used in handheld terminals with small resources.

The proposed algorithm provides VBR bit streams by controlling QP on a per picture basis. The QP is calculated based on two other QPs, which conceptually correspond to the CBR and constant quality rate controls. The QP corresponding to the CBR is computed based on short-term variations in the bit rate and the QP corresponding to the constant quality is computed based on long-term variations in the bit rate. With this structure the proposed RCA can be tuned to operate at any point between CBR and constant QP.

The algorithm utilizes the variable bit rate benefits to minimize the unnecessary variations of QP and to provide encoded video with high visual quality. The algorithm allocates a bit budget to a *Set Of Frames* (SOF) based on the buffer state and a heuristically optimized bit allocation function. Allocated bits to the frames in the SOF are computed based on the SOF bit budget and the picture types including I, P, and B types. The buffer state is not used during the encoding of a SOF to prevent unnecessary variations in encoding parameters. After bit allocation, the two virtual QPs are computed. A global QP is computed for each SOF that corresponds to a long-term rate control and a local QP is computed for each frame that corresponds to a CBR rate control. The final real QP is computed on a frame basis using the virtual QPs. The values of the local QP and global QP can be close or far from each other. The local QP is smoothed by a simple low-pass filter and then, a nonlinear modulation on the local and the global QP is implemented to compute the final QP. The low-pass filter and modulation function have been designed such that the overall variations in the QP are minimized.

To decrease the computational complexity, a simple first-order RD model for the calculation of local and global QPs is used. Moreover, the coding complexities of pictures are estimated based on the coding complexities of previous encoded pictures. This significantly reduces the computational complexity of the algorithm.

The proposed real-time variable rate control algorithm has been implemented in H.263 and MPEG-4 (Part 2) encoders. The performance of the proposed RCA was compared with the RCA presented in H.263 TMN8 [Gardos 1997] and also with the two-pass RCA presented in [Varsa 2001]. The algorithm performs significantly better than the mentioned anchors on a large set of known video sequences. When comparing to the two-pass algorithm, it provides 0.98 dB enhancement in average quality, 42% decrease in real buffer size used in streaming, and 20% decrease in required delay in streaming. When compared to TMN8 RCA, it results in 0.42 dB improvement in average quality. While this improvement is not very spectacular, it should be kept in mind that TMN8 drops a large number of frames, about 4%, whereas the proposed algorithm was configured not to drop any frame. Moreover, the average QP in our algorithm is 0.78 lower than in TMN8 RCA. Consequently, the algorithm is expected to provide a much better visual quality than that of TMN8. According to the experimental results, the algorithm strictly obeys the buffering constraint while it can be tuned to operate on a wide range of buffering delays. Moreover, it can be used for different hybrid video coding standards such as H.263 and MPEG-4.

2.2. RD MODEL FOR STANDARD VIDEO CODECS

A Rate-distortion model is the basic parts of many practical RCAs. The RD model is used to compute the QP based on the allocated bit budget and according to the estimated coding complexity for a MB or frame. In the proposed RCA in [P1] a simple first-order RD model was used for the calculation of QP and the coding complexity of frames was estimated based on previous encoded frames. There is a problem in the calculation of QP for the first intra pictures in the sequence and also for pictures at scene cut boundaries. A video sequence may have very different scenes and the complexity estimation based on previous frames or scenes may be inaccurate. Moreover, using intra pictures at scene cuts provides a better RD performance. Therefore, it would be beneficial to compute an accurate QP for an intra picture based on a complexity measure that is calculated from the picture itself. The research work in this direction continued and lead to publication [P2].

In [P2] a general and precise RD model for standard hybrid video codecs, including H.263, MPEG-4, and H.264/AVC is proposed. It is designed taking into account previous theoretical results, and based on new assumptions which are confirmed by comprehensive practical experiments. The proposed RD model is divided conceptually into two separate parts. The first part is codec specific and reflects the codec's properties. The second part is video content specific and describes the interaction between the video content and the encoder. The general form of model can be expressed as

$$R(Q) = E(Q^{-1}) + F(Q^{-1}) \times X \,, \tag{2.1}$$

where E and F are two function of QP as the codec specific and the content specific parts of the model, respectively. X denotes the coding complexity and R stands for the estimated bit budget used for encoding when the QP of Q is used. The proposed RD model is a generalized form of several previously used models in which the codec specific part of the model has been ignored or simplified to a constant value independent of QP. See the RD models in [Chiang 1995], [Chiang 1997], [Gardos 1997], [Lee 2000], [ISO/IEC 2001] as examples. Considering the codec specific part of the RD model as a function of QP provides more accuracy for the model while it can be parameterized only once for each codec and can be used henceforth with no need for updates during encoding. The model can be parameterized for all types of frames and complexity measures. It can be parameterized by the MAD of motion compensated blocks in inter-prediction frames. It can also be parameterized for intra frames by a complexity measure such as the variance which is computed based on the frame.

A new measure for coding complexity of video content is also introduced in [P2] that specifies the relationship between rate and the video content properties. Different complexity measures have been used in previous RCAs. Generally, the number of coded bits generated in a block is proportional to the block variance. In TMN8 RCA [Gardos 1997], the rate of an intra or motion-compensated MB is modeled as a function of variance of luminance and chrominance values in the MB. In MPEG-4, Appendix-L [ISO/IEC 2001] the blocks are classified according to their variance and then a look up table is used for the estimation of encoded bits. In some RCAs, such as in [Lee 2000], [Ma 2003], and [Sullivan 2003], the MAD of motion-compensated MBs is used as a measure of complexity. From a computational complexity point of view, the MAD is preferred because it is calculated for the motion estimation process. Also the *Sum Absolute Difference* (SAD) for motion-compensated MBs is used as a coding complexity measure in some RCAs such as [Ngan 2003] and [Ma 2005]. MAD and SAD are appropriate for B and P frames but cannot be used for intra frames. The proposed coding complexity measure in [P2] not only can be computed over intra pictures or MBs but it can also be computed over motion-compensated MBs.

Among the known complexity measures, the variance seems to give the most accurate performance in many cases. However, experimental results show that it is insufficient when some special textural structures exist in the video. A video frame with special textures may need a large amount of bits for encoding while it has a small variance. On the other hand, some textural structures with a large variance may need only a few bits for encoding.

To reflect the shortcomings discussed above, we introduce the following coding complexity measure:

$$X = \overline{V} + \overline{T_V} + \overline{T_H} \ , \tag{2.2}$$

where

$$V = \frac{1}{256} \sum_{i=1}^{16} \sum_{j=1}^{16} \left(P(i,j) - \overline{P(i,j)} \right)^2 , \tag{2.3}$$

$$T_V = \sum_{i=1}^{16} \sum_{j=1}^{16} \left| P(i,j) - P(i,j-1) \right| , \tag{2.4}$$

$$T_H = \sum_{i=1}^{16} \sum_{j=1}^{16} \left| P(i,j) - P(i-1,j) \right| , \tag{2.5}$$

where X denotes the complexity measure. V is the variance of luminance pixels $P(i,j)$ in one MB, T_V and T_H denote two vertical and horizontal texture measures on the luminance pixels. $\overline{V}$, $\overline{T_V}$ and $\overline{T_H}$ are the average values of V, T_V and T_H, respectively, on a number of MBs in the frame. Experimental results indicate that these parameters can be independent, and each of them plays an important role in coding complexity. Assuming an equal degree of relevance, they get equal weighting factors in the final closed form formula (2.2).

A number of simulations were performed on three standard video codecs including H.263, MPEG-4 (part 2), and H.264/AVC to evaluate the performance of the proposed RD model and also the proposed coding complexity measure. According to the simulation results, the proposed RD model and complexity measure always outperform the best known competitors. As typical results when the proposed RD model is parameterized by the proposed complexity measure, it can estimate the required bits for encoding an intra picture with an estimation error about 10%. This provides a very good tool for the rate control to compute a precise QP for encoding intra frames.

2.3. BIT ALLOCATION ALGORITHM FOR VBR VIDEO

In [P3] a bit allocation method for the VBR video is proposed. The idea is to make some intentional variations in the bit rate of the encoded bit stream in order to increase the overall quality. A *Special P* (SPP) frame is introduced to implement the proposed bit allocation method. The SPP frame is a high quality inter-prediction P picture that is inserted according to some conditions into the bit streams.

A simple block diagram for encoding P-frames is shown in Figure. 2. In this diagram, P(n+1) denotes an uncompressed P-frame and P(n) denotes the uncompressed P-frame before P(n+1). P(n+1)+E(n+1) and P(n)+E(n) represent the reconstructed version of the P(n+1) and P(n) frames, respectively. E(n+1) and E(n) are the quantization errors in these frames. C(n) denotes the motion compensated frame used for encoding P(n+1). If the QP used for encoding R(n+1) is greater than the QP used for R(n), then likely E(n+1)>E(n). The higher QP distributes the quantization error over more frequency components. Therefore, the signal R(n)+E(n) has some high frequency information which is not available in the signal R(n+1)+E(n+1). This high frequency information will appear in C(n) and then in P(n+1)+E(n+1) which is the reconstructed version of P(n+1). While consequent frames have common information in all frequency components, it is likely that the quantization error of E(n+1) decreases by the injection of high frequency information available in the previous frame. The SPP frames utilize this effect. This effect has been used in a different way by Pan *et al.* in [Pan 2003] and Chellappa *et al.* in [Chellappa 2004]. Pan *et al.* proposed a bit allocation algorithm in which allocated bits to a P frame in a GOP decrease proportionally to its distance from the end of GOP. Chellappa *et al.* proposed a dual frame motion compensation method in which one short-term reference frame and one long-term reference frame are used for motion compensation to allocate bits unevenly among frames. This method periodically creates high quality frames that serve as long-term reference frames.

The SPP frame is a P-frame with a relatively small QP and high bit rate in special locations in the bit stream. The QP and the location of SPP frames are very important parameters which should be selected carefully. The aim is to take advantage of SPP frames without any cost on the average bit rate and transmission delay. When a bit budget larger than normal is allocated to a SPP frame, the bit budget allocated to the subsequent frames should decrease to provide a long term constant rate. The quality of the subsequent frames may degrade as a result. The idea is to find an optimal way to relocate bit budget from subsequent frames to SPP frame. A method for implementing SPP frames that can be used in conjunction with any VBR video RCA that operates on picture basis is proposed.

According to the algorithm, the RCA defines a QP for encoding each video frame. A control block placed between RCA and the encoder changes the computed QP by RCA properly to implement the proposed bit allocation algorithm. QP of the SPP frame is computed based on the QP of the normal P-frame at the same location which is computed by RCA as:

$$Q_{SPP} = Q_P \times (\Delta + \lambda \times Q_{SI}^A / Q_P^A), \tag{2.6}$$

where Q_P is the value of QP calculated by the rate control for the normal P-frame independently of the SPP frame. Q_P^A denotes the average of QP over all encoded P frames and Q_{SI}^A stands for the average of QP over all P-frames from the last SPP or intra frame to the current frame. Δ and λ are two constant coefficients smaller than one which are defined heuristically. In this equation, two assumptions have been considered. First, the QP

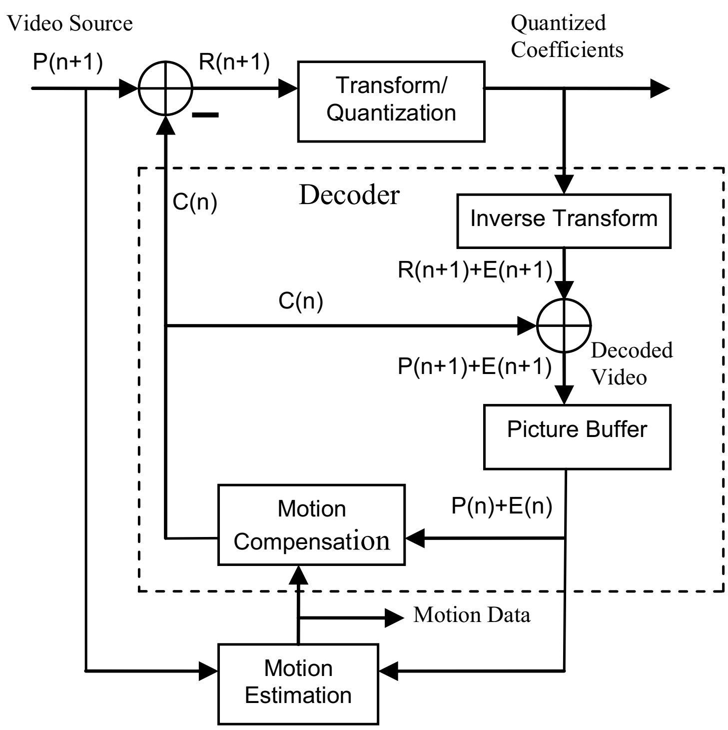

Figure 2: Simple diagram for encoding of P-frames

of SPP frame is proportional to that of the QP of the normal P frame. Second, when Q_{SI}^{A} has a large value relative to Q_{P}^{A}, it means the changes between adjacent frames are higher than average. In this case, the effect of SPP-frame becomes limited to a smaller number of subsequent frames; and, therefore, a relatively smaller bit budget is allocated to the SPP frame. This is executed by increasing the QP of SPP frame according to the local average of QP.

The SPP frames do not have a constant frequency or interval and they are inserted in a video stream according to certain conditions. Due to the application and the RCA, different conditions can be considered. The following essential conditions are considered in [P3]:

1) Min SPP Interval: SPP interval defines the distance between two SPP frames or between one intra frame and the next SPP frame. As a condition, the SPP interval should be greater than or equal to a threshold value $N(\min)$ as

$$N(\min) \leq SPP \quad Interval. \tag{2.7}$$

The minimum value of SPP interval is defined according to the local and global average values of QP as

$$N(\min) = \phi - \theta \times Q_{SI}^{A} / Q_{P}^{A}, \tag{2.8}$$

where ϕ and θ are two constant coefficients. When the local QP is large, it means the changes between adjacent frames are high. In this case, the effect of SPP-frame is limited to a smaller number of subsequent frames. Therefore, the SPP frames should be closer together.

2) Buffer State: When a RCA with buffering constraint is used, the buffer state can imposed some conditions on the SPP frames. Considering a buffering model corresponding to the receiver side for the buffer, the following two conditions are proposed. A SPP frame can be inserted if:

$$B > \left(\; Min(B) + \alpha \times R_P \; \right), \tag{2.9}$$

$$B > B_S \times MFL, \tag{2.10}$$

where B and B_S denote the buffer fullness and buffer size, respectively. R_P shows the allocated bits to the normal P-frame computed by the rate. α and MFL (Margin Factor for Low fullness) are two constant values. The condition defined in (2.9) prevents inserting an SPP frame if it would lead to an unnecessarily streaming delay. Condition (2.10) prevents using an SPP frame when the buffer is close to under flow. Similar conditions can be used according to the application.

It is shown in [P3] that the changes of QP on the P-frames followed by the SPP frame are negligible. Therefore, the SPP control block just tunes the location and the QP of the SPP frame without any interaction with the rate controller. After encoding an SPP frame, the QP of subsequent frames can be tuned by the RCA that is normally operating. Hence the proposed SPP frame can be implemented independently of the RCA and this is one of the main advantages of the proposed bit allocation method.

The proposed bit allocation method has been implemented and evaluated on H.263 and MPEG-4 (part 2) encoders, baseline profile, and single reference frame. Evaluation results on a large number of known video sequences show that the SPP frames improve the average PSNR (0.32 dB in H.263 and 0.24 dB in MPEG-4) of encoded bit streams without any changes in average bit rate and buffering delay.

2.4. REGION BASED MINIMIZATION OF QUANTIZATION ERROR PROPAGATION IN VBR VIDEO

In [P3], a special bit allocation algorithm for VBR video was presented. The idea of SPP frames was introduced to implement the proposed bit allocation method. In fact the SPP frames improve the average quality of decoded video by decreasing propagation of quantization errors in subsequent frames. A special amount of details is carried once by an SPP frame which will appear in several subsequent frames. As a result, the SPP frame increases the average quality of decoded video. According to the simulation results, the performance of SPP frames depends on the amount of motion and motion properties in the video sequence.

In [P4], the idea of SPP frame is extended to MB level to minimize the propagation of quantization error based on motion properties in different regions of video frames. The video frames are divided into different regions based on motion properties in the regions. Depending on the motion prediction process, the propagation of quantization error is different in each region. In static regions or in regions with relative simple motions, the quantization error may propagate just temporarily while in regions with complex motions, the quantization error is propagated temporarily and spatially too. The content of error also depends on the motion properties. The quantization error in a static region or in a region with simple motion, in which motion prediction compensation is near perfect, can propagate to the next frames without any change. On the other hand, in a region with complex motion, the error changes between frames because of defective motion prediction compensation and integration to new errors.

Special P macroblock (SPPMB) is introduced to cope with the different types of quantization errors. An SPPMB is an intra/P MB with special temporal and spatial location and a relative large bit budget that reduces the propagation of quantization errors to the following frames. Where a near perfect motion prediction compensation is possible, SPPMBs carry high frequency information to the decoder side which increase the quality of several subsequent frames without any cost in bit budget. The temporal and spatial locations of SPPMBs and their bit budgets are very important parameters which should be determined carefully in order to increase the average quality of VBR video without any extra cost in average bit rate and delay.

A simple method for implementation of SPPMB is proposed in which a control block placed between the rate controller and the encoder defines the locations and the QPs of SPPMBs independently of the RCA. Figure 3 shows a simple block diagram for the implementation of SPPMB. The rate controller operates at the frame level and it computes a QP for each frame. The SPPMB control block uses the QP computed by the rate controller as input and determines the QP and the locations of SPPMB in the frame. Any VBR video RCA with control on frame level such as our presented controller in [P1] can be used in this structure. The SPPMB control block also uses a region map which defines the different regions in the video frames. The region map defines two or more types of regions according to motion properties in the frame. The region is defined as a group of connected neighboring MBs which have similar motion properties in a number of consequent frames.

The QP of SPPMB, it is calculated from the QP of the whole frame which is computed by the rate controller:

$$Q_S = Q_P \times (\Delta + \lambda \times Q_{SI}^A / Q_P^A),\ Q_S \le Q_P \ , \tag{2.11}$$

where Q_S denotes the QP of SPPMB and Q_P is the value of QP calculated by the rate control for the P-frame independently of the SPPMBs. Q_{SI}^A denotes the average of QP over all corresponding P MBs from the last SPPMB or intra MB in the previous frames to the current MB. Q_P^A denotes the average of QP over all encoded corresponding P MBs. Δ and λ are two constant coefficients which are selected according to a region map from a set of experimentally optimized parameters. In definition of (2.11) two assumptions are considered. First, the QP of SPPMB is proportional to QP of the normal P MB at the same location. Second, when the local average Q_P^{LA} has a large value relative to the global average Q_P^A, it means a larger prediction error relative to the average over all encoded frames. In this case, the effect of SPPMB becomes limited to a smaller number of subsequent frames and therefore, a relatively smaller bit budget is allocated to the SPPMB. This is implemented by increasing QP of SPPMB according to the ratio of the local average of QP to the global

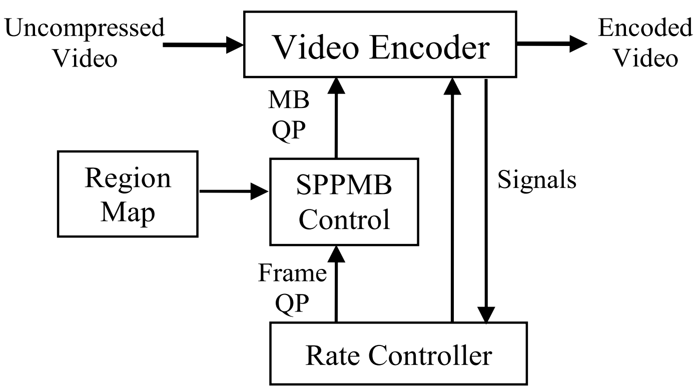

Figure 3: Diagram for implementation of SPPMB

average of QP. The constant coefficients used here depend on the encoder and the region map.

SPPMBs do not have a constant temporal frequency or interval. A normal inter/P MB is replaced with a SPPMB if a set of conditions are met. Although the conditions may be modified according to the application, two essential conditions are considered in [P4]. First, similar to the SPP frames a minimum interval is defined for the SPPMBs. Second, the total number of SPPMBs in one frame is limited to an upper bound to distribute the extra bit budget of SPPMBs between several frames and to prevent P frames with a large bit budget.

After encoding a P frame with a number of SPPMBs, the QP of subsequent frames is computed by the rate controller and without any interaction with the SPPMB control block. Hence, the proposed SPPMBs can be implemented independently of the RCA and this is one of the main advantages of the proposed method.

The proposed method has been implemented and evaluated on H.264/AVC codec, baseline profile using single reference frames. Simulation results on a set of known video sequences show that the proposed SPPMBs can improve the average quality (about 0.71 dB in PSNR) of compressed video. In comparison to the SPP frames, SPPMBs provide a similar enhancement in PSNR but they considerably (about 1.8) decrease the average QP used for encoding the regions that have complex motions. This means a higher visual quality in the regions that have complex motion.

More experiments on the region map resolution show that a simple binary map with only two types of regions is enough for many video sequences. The proposed bit allocation method can be used *in Region of Interest* (ROI) video coding where the region map may or may not by correlated with the ROI.

2.5. A SEMI-FUZZY RCA FOR VIDEO STREAMING

2.5.1. Fuzzy Control Definitions and Terminology

Fuzzy control is one of the most successful applications of *fuzzy logic*. Fuzzy logic was introduced by Zadeh in 1965 [Zadeh 1965] as an extension of *Boolean logic* to model the uncertainty of natural language. Fuzzy control mimics human control logic. It uses an imprecise but very descriptive language to operate like a human operator. Fuzzy control does not require a mathematical model for the system under control and it can be simply implemented for very complex nonlinear systems. In fuzzy control, linguistic information from human experts is expressed by fuzzy *IF-THEN* rules. The fuzzy rules can describe the control actions that should be taken (direct control approach) or they may describe the behaviors of the system under control (indirect control approach) [Wang 1993]. Some definitions and terminology used in the fuzzy logic control are listed below.

Crisp Sets and Fuzzy Sets

Normal sets in classic set theory are called *crisp* sets as compared to fuzzy sets in fuzzy set theory. Let *C* be a crisp set defined on the universe of discourse *X*, then for any element *x* of *X*, either $x \in C$ or $x \notin C$. In a fuzzy set *F* it is not necessary that either $x \in F$ or $x \notin F$. A *membership function* (MSF) assigns to each elements *x* in *X* a value from the unit interval [0, 1] that shows the degree of membership. A fuzzy set *A* in the universe *X* is defined by $A = \{(x, \mu_A(x)) \mid x \in X\}$, where $\mu_A(x) \in [0,1]$ is the membership function of *x* in *A*.

Fuzzy Singleton

A fuzzy set $A = \{(x, \mu_A(x)) \mid x \in X\}$ is said to be a *fuzzy singleton* if $\mu_A(x) = 1$ for $x \in X$ and $\mu_A(x) = 0$ for all $x' \notin X$ with $x' \neq x$.

Cartesian Product

If *X* and *Y* are two universal sets, then $X \times Y$ is the set of all ordered pairs (*x*, *y*) for $x \in X$ and $y \in Y$. Let *A* be a fuzzy set of *X* and *B* be a fuzzy set of *Y*. The Cartesian product is defined as $A \times B = \{(z, \mu_{A\times B}(z)) \mid z = (x, y) \in Z, Z = X \times Y\}$, where $\mu_{A\times B}(z) = \mu_A(x) \wedge \mu_B(y)$.

Fuzzy Relation

Fuzzy relation is used to describe the association between two things. If *R* is a subset of $X \times Y$, then *R* is said to be a relation between *X* and *Y*, or a relation on $X \times Y$. A fuzzy relation is a fuzzy set as $R(x, y) = \{((x, y), \mu_R(x, y)) \mid (x, y) \in X \times Y, \mu_R(x, y) \in [0,1]\}$, where $\mu_R(x, y)$ is the membership function of *R*.

Composition of Fuzzy Relations

The composition operations provide means for combining several fuzzy sets and relations. Let R and S be fuzzy relations on $X \times Y$ and $Y \times Z$ respectively. The *so-called Sub-Star composition* of *R* and *S* is a fuzzy relation on $X \times Z$ defined as:

$$R \circ S \Leftrightarrow \mu_{R\circ S}(x, z) = \mathrm{sub}_{\mathrm{y}\in\mathrm{Y}} \{\mu_R(x, y) * \mu_S(y, z)\},$$

where the "$*$" can be any t-norm operator. The *t-norm* operator corresponds to the conjunction "AND" and the most commonly used operations for t-norm are

$$T_m(x, y) = \min(x, y), \quad \text{Fuzzy Intersection}$$

$$T_b(x, y) = xy, \quad \text{Algebric Product}$$

$$T_p(x, y) = \max(0, x + y - 1), \quad \text{Bounded Product}$$

Fuzzy Implication

Fuzzy implication is used to represent fuzzy rules. It is a mapping from an input fuzzy region A onto an output fuzzy region B according to a defined fuzzy relation R on $A \times B$ as $B = A \circ R$. It is denoted by $A \to B$.

Fuzzy Logic Control System

A basic block diagram of a fuzzy logic controller is shown in Figure 4. The inputs to the controller are feedback crisp signals from the output of the system under control and the outputs of fuzzy controller are the crisp control signals.The fuzzy controller performs a mapping from crisp inputs to crisp outputs. The *fuzzifier* maps a crisp point at the input into a fuzzy set. As example a *singleton fizzifier* maps a crisp input into a fuzzy singleton. The *fuzzy rule base* consists of a collection of fuzzy IF-THEN rules. Mamdani fuzzy rules and Taksgi-Sugeno (TS) fuzzy rules are two popular types of fuzzy rules. A general form of Mamdani fuzzy rule can be expressed as:

$$R_M^l : \text{ IF } x_1 \text{ is } F_1^l \text{ AND ... AND } x_n \text{ is } F_n^l \text{ THEN } y \text{ is } G^l, \quad l = 1, ..., m$$

where $x = (x_1, ..., x_n)^T \in X$ and $y \in Y$ are the input and output linguistic variables respectively. F_i^l and G^l are labels of fuzzy sets in X_i and Y respectively, where $X = X_1 \times ... \times X_n$.

The general form of a TS fuzzy rule is described as:

$$R_T^l : \text{ IF } x_1 \text{ is } F_1^l \text{ AND ... AND } x_n \text{ is } F_n^l \text{ THEN } y^l = c_0^l + c_1^l x_1 + ... + c_n^l x_n, \ l = 1, ..., m$$

where F_i^l are fuzzy sets, c_i^l are real valued parameters, y^l is input of the system under control. In the TS fuzzy rules, the premise is fuzzy but the consequence is crisp.

The *fuzzy inference engine* performs a mapping from fuzzy sets in $X = X_1 \times ... \times X_n$ to fuzzy sets in Y. The fuzzy inference engine uses *fuzzy reasoning* to conclude a fuzzy set for each fuzzy rule according to the input information. The fuzzy reasoning based on the *generalized modus ponens* (GMP) can be written in IF-THEN form as

Premise 1:	IF x is A THEN y is B
Premise 2:	x is A'
Consequence:	y is B'

where A, A', B, and B' are fuzzy sets. Using compositional rule of inference B' can be computed as

$$B' = A' \circ R = A' \circ (A \times B).$$

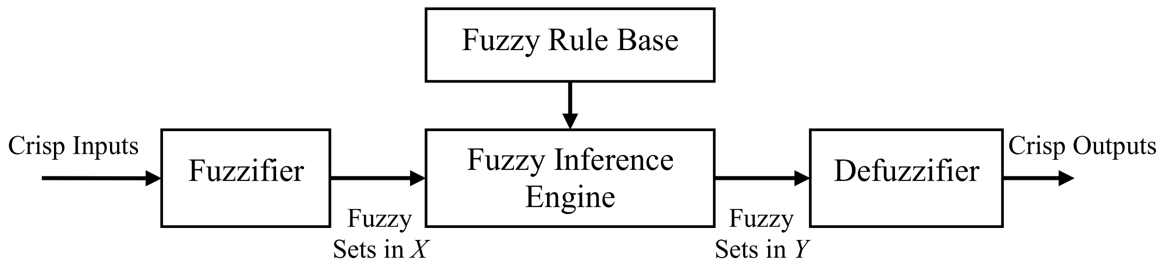

Figure 4: A basic diagram of a fuzzy logic controller

The final fuzzy set is obtained by combining the fuzzy sets concluded from each fuzzy rule. The *t-conorm* operators that correspond to the fuzzy disjunction "OR" are used for combining fuzzy sets. The most commonly used operations for the t-conorm are

$$C_m(x, y) = \max(x, y), \qquad \text{Fuzzy Union}$$

$$C_b(x, y) = \min(1, x + y), \qquad \text{Bounded Sum}$$

$$T_p(x, y) = x + y - xy, \qquad \text{Algebric Sum}$$

The *defuzzifier* maps the fuzzy sets in Y to a real number. The defuzzified output is treated as the control signal for the process under control. Popular defuzzification methods include the *center-average* (*centroid) defuzzifie*r, and the *mean-of-maxima defuzzifier*. The center-average defuzzifier generates a control signal by finding the center of gravity of the area surrounded by the aggregated membership function and horizontal axis. The mean-of maxima defuzzifier generates a control signal that represents the mean value over all fuzzy sets whose membership functions reach the maximum. A review on fuzzy logic controllers is presented in [Lee 1990].

Fuzzy Controllers

In this thesis, fuzzy controllers have been utilized in different scenarios. A simple fuzzy system with two inputs including singleton fuzzifier, center-average defuzzifier, and *product inference engine* in which the algebraic product is used for t-norm operations, have been employed in all the scenarios. However, the structure of the system under control, the fuzzy rules, and fuzzy membership functions have been designed properly to the scenarios.

2.5.2. Semi-Fuzzy Rate Control Algorithm

A RCA for real-time VBR video application was proposed in [P1]. It performs well but it has many tuning parameters that should be adjusted according to the application. A fuzzy-logic-based RCA for MPEG video is presented in [Tsang 1998]. We proposed two other fuzzy RCAs for mobile and streaming applications in [S1][S2]. The presented RCA in [S1] is a simple fuzzy rate controller for mobile applications including local recording, streaming, and multimedia messaging services (MMS). This RCA has been implemented on MPEG-4 (part-2) and H.263 and good results were obtained. The proposed RCA in [S2] is a fuzzy controller which is optimized for video streaming applications with a relatively high delay. It also utilizes the coding complexity measure proposed in [P2] to calculate the QP of intra pictures. Another fuzzy RCA is proposed with delay constraint optimized for video streaming over DVB-H applications in [S3]. This RCA provides more control over the bit rate in a low delay. A generalized form of this RCA has been published in [P5].

The proposed RCA in [P5] is a *Semi-Fuzzy* (SF) controller that can be used for a wide range of VBR applications. The term semi-fuzzy is adopted because a fuzzy controller cooperates with a classic controller. The semi-fuzzy RCA is a modified version of our previous RCA that is tuned easily for different applications. The input signals, fuzzy rules, and fuzzy membership functions have been modified. Furthermore, in the new version, the gain of the fuzzy feedback loop is adjusted adaptively according to the application. Moreover, the QP of I-pictures is computed adaptively to the application and content. The adaptation features used in the new RCA makes it easy to be tuned for different applications.

The proposed algorithm has a low degree of complexity. It does not use any RD model directly and therefore, there is no need to update the RD model parameters based on a complexity measure such as variance or MAD. The proposed RCA can be used for almost all hybrid video coding standards with any rate distortion optimization process [Sullivan 1998], [Wiegand 2001], [Wiegand 2003a]. There is no chicken-and-egg dilemma for H.264/AVC encoder anymore because the QP is calculated only based on results of previous encoded pictures.

Overview of RCA

The semi-fuzzy RCA controls the bit rate by adjusting the QP on a picture basis. It utilizes a fuzzy rate controller in conjunction with a quality controller and several other tools to calculate the QP for different types of video picture. Although here only intra-prediction pictures (I-picture) and reference inter-prediction pictures (P-pictures) are explained, the algorithm can be easily extended to other types of pictures as well. The fuzzy controller utilizes a virtual buffer to impose buffering constraints on the bit stream. The quality controller uses another feedback from the PSNR of encoded video to minimize the variations of picture quality.

The RCA can be functionally divided into two main parts. The first part utilizes the fuzzy controller and the quality controller to compute the QP of P-pictures. The second part of algorithm uses several other feedback signals from the buffer state, uncompressed and compressed video to calculate the QP of I-Picture based on some functions derived from a simple RD model. The I-pictures at the scene cut boundaries are treated differently from the normal I-pictures at the periodic random access points. In VBR, the bit allocation to I-pictures has a remarkable impact on the overall RD performance. Therefore, the QP of I-pictures is computed very carefully in this RCA. The key point in the proposed RCA is to prevent unnecessary variations in quality while the buffer constraint is obeyed. Some details about the semi-fuzzy RCA are presented in the sequel.

Reference Inter Prediction Pictures

The QP for P-pictures is defined by the fuzzy controller and the quality controller. Figure 5 depicts the block diagram of the proposed rate control system for the P-pictures. The fuzzy controller, the quality controller and the virtual buffer are the basic elements of the control system. The fuzzy controller attempts to control the bit rate of the encoded bit stream by controlling the variation of QP while it has been optimized such that to prevent unnecessary fluctuations of QP. In computation of QP, it is assumed that the subsequent video pictures have a similar degree of complexity (except in scene cuts) so the complexity of the previous encoded picture is used as an estimate for the complexity of the subsequent picture and the QP of the subsequent picture is computed based on QP of previous encoded picture with small variation which is defined by both the fuzzy and the quality controllers. The fuzzy controller uses two feedback signals from the buffer fullness and the bit rate. The quality controller utilizes a feedback signal from the quality (PSNR) of encoded. Furthermore, a low pass filter (LPF) smoothes the feedback signal from the rate to the fuzzy controller to smooth the variations in the output of the fuzzy controller. The QP for the current P-picture (Q_P) is the sum of the QP used for encoding the previous picture and the output of the fuzzy controller (ΔQ_F) added to the output of the quality controller (ΔQ_Q) or

$$Q_P(i) = Q_P(i-1) + \Delta Q_F(i) + \Delta Q_Q(i)\,. \tag{2.12}$$

From the system point of view, the main part of the computed QP for a P-picture is the delayed version of QP used for the previous picture and the control (variation) of QP is provided by the fuzzy controller and the quality controller. From the RD performance point of view, the RD performance of previous encoded pictures is used as reference for the next picture and small deviation from the reference point is computed. The main advantage of this approach is that in the small range around the reference point, all nonlinear functions that exist in the system can be assumed as linear without losing computational accuracy. More details about the RCA elements are presented below.

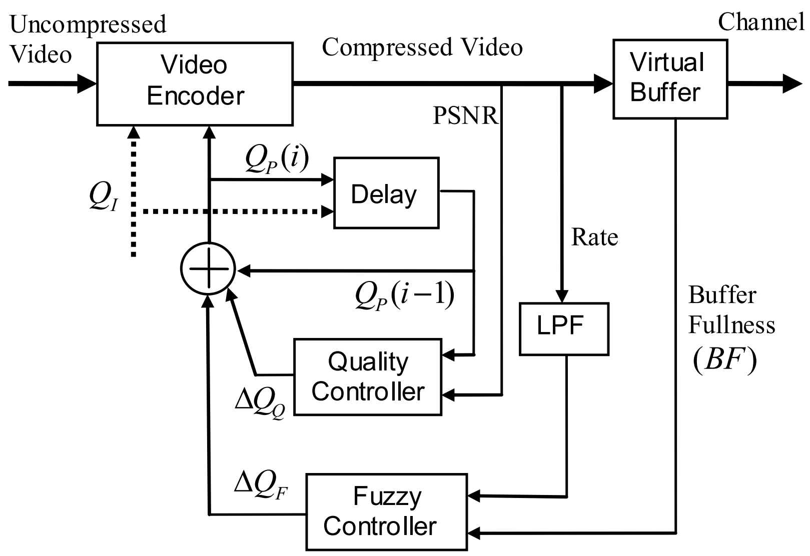

Figure 5: Block diagram of the RCA for P-Pictures

Virtual Buffer

The virtual buffer used by the controller simulates the buffering process of the decoder in the receiving side of a CBR channel. Although it utilizes a simple model, it is nearly identical to the hypothetical reference decoder models used in different video coding standards. The occupancy of virtual buffer is updated after encoding each video picture as:

$$O_B(i+1) = O_B(i) - B(i) + (R_T / F), \tag{2.13}$$

Where $O_B(i)$ denotes the occupancy of virtual buffer before encoding the i^{th} picture. $B(i)$ shows the number of bits consumed by the i^{th} encoded picture (P or I). R_T indicates the target average bit rate for the bit stream or the channel bandwidth and F stands for the frame rate. Note that the virtual buffer models the decoder buffer at the receiver side. Therefore, the occupancy of this buffer corresponds to the free space of a buffer at the encoder or transmitter side.

Fuzzy Controller

From the non-fuzzy rate control approach we learned that many heuristic functions coexist with the nonlinear RD models in the rate control process. As a new approach, the fuzzy controller is selected for this structure because nonlinear functions and the complexities that exist in rate control tasks can be simply included in the fuzzy rules and fuzzy membership functions. Generally, a fuzzy controller can be designed based on the expert experiences or it can be learnt from examples. Therefore, a fuzzy controller is a good option to use the many heuristic results for the video rate control. Moreover, according to the block diagram shown in Figure 5, a controller is required to define a small quantized value based on rough measurements on the bit rate and buffer fullness. These properties make it fit to a fuzzy controller with low resolution inputs and output.

The fuzzy controller has two input signals that are normalized values of the buffer occupancy and the actual bit rate of P-pictures. Buffer occupancy is normalized by the buffer size and the actual bit rate of P-pictures is normalized by the target bit rate for P-pictures. While in VBR the consumed bit budget by P-pictures can be very different from that of I-pictures, depending on the frequency of I-pictures in the bit stream, the target bit rate of P-pictures can be very different from the whole target bit rate. It is attempted to estimate a precise value for the target bit rate of P-pictures to be used for the normalization purpose. The fuzzy inputs are defined as:

$$x_1 = O_B / S_B , \tag{2.14}$$

$$x_2 = \frac{B_P F}{R_T}\left(1 + \frac{X_{IP} - 1}{I_I}\right), \tag{2.15}$$

where B_P denotes the consumed bit budget by the previous encoded P-picture. I_I stands for the interval of periodic I-pictures in the bit stream in terms of number of pictures. X_{IP} indicates the coding complexity of I-pictures relative to P-pictures and it is computed as:

$$X_{IP} = \overline{B}_I / \overline{B}_P , \tag{2.16}$$

where $\overline{B}_I$ and $\overline{B}_P$ denote the average consumed bit budgets by the encoded I-pictures and P-pictures, respectively. If the previous encoded picture is an intra picture, the value of B_P in (2.15) is reset to the value of B_I / X_{IP}. To suppress the fluctuations of QP results of short-term variations in complexity of video pictures, the low pass filter (LPF) smoothes the variation of B_P before input to the fuzzy controller. The impulse response of used LPF is:

$$H(z) = m / \left(m + 1 - z^{-1}\right), \tag{2.17}$$

where m is a constant value and good results are obtained with m=1.2.

All fuzzy rules and the output of the fuzzy controller are summarized in Table I. The letters H, L, M and V correspond to the fuzzy descriptions High, Low, Medium, and Very, respectively. The number before V shows the number of repetition or level of strength. As an example from the table, it can be expressed as: *if* (x_1 is VL and x_2 is H) *Then* (Output is 3VH). 3VH stands for Very Very Very High. The input signals are specified by their fuzzy membership functions. The inputs x_1 and x_2 employ nine and seven fuzzy membership functions, respectively. The fuzzy rules and membership functions were designed based on the experiences form our previous RCA. The asymmetric structures in the table of fuzzy rule and fuzzy MSF are related to a number of facts which affect the operation of RCA. The nonlinearity of the RD function and the difference between the bit budgets of I and P-pictures are two key points that cause the asymmetry in the structures. The other key point is that the control loop gain is a function of buffer conditions. A more aggressive control is required when the buffer fullness is close to critical conditions to prevent underflow and overflow and a looser control is preferred when the buffer fullness is far from the critical conditions to prevent unnecessary variations in quality of encoded video. After preliminary design of the fuzzy system, an optimization process was performed for fine tuning the fuzzy membership functions. In the optimization process several parameters including average bit rate, average PSNR, average QP, and standard deviation of PSNR were considered. The final membership functions are depicted in Figure 6. The desired central values for the output of the fuzzy system correspond to the fuzzy rules in Table I are depicted in Table II.

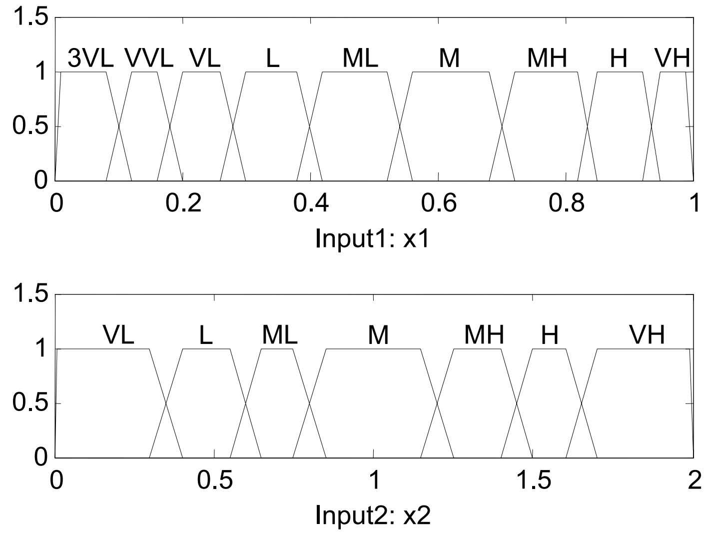

Figure 6: Membership functions of the linguistic variables

TABLE 1: SUMMARY OF THE IF-THEN FUZZY RULES

x_2 \ x_1	XL	VVL	VL	L	ML	M	MH	H	VH
VH	6VH	5VH	4VH	3VH	VVH	VH	H	MH	M
H	5VH	4VH	3VH	VVH	VH	H	MH	M	ML
MH	4VH	3VH	VVH	VH	H	MH	M	ML	L
M	3VH	VVH	VH	H	MH	M	ML	L	VL
ML	VVH	VH	H	MH	M	ML	L	VL	VVL
L	VH	H	MH	M	ML	L	VL	VVL	3VL
VL	H	MH	M	ML	L	VL	VVL	3VL	4VL

TABLE 2: DESIRED CENTRAL VALUES FOR THE OUTPUT OF THE FUZZY SYSTEM

x_2									
VH	8	7	6	5	4	3	2	1	0
H	7	6	5	4	3	2	1	0	-1
MH	6	5	4	3	2	1	0	-1	-2
M	5	4	3	2	1	0	-1	-2	-3
ML	4	3	2	1	0	-1	-2	-3	-4
L	3	2	1	0	-1	-2	-3	-4	-5
VL	2	1	0	-1	-2	-3	-4	-5	-6
x_1	**3VL**	**VVL**	**VL**	**L**	**ML**	**M**	**MH**	**H**	**VH**

A well-known and simple fuzzy system with two inputs using product inference engine, singleton fuzzifier, and center-average defuzzifier, was used

$$f(x_1, x_2) = \frac{\sum_{i_1=1}^{N_1}\sum_{i_2=1}^{N_2} \bar{y}^{i_1 i_2} \mu_{A_1^{i_1}}(x_1).\mu_{A_2^{i_2}}(x_2)}{\sum_{i_1=1}^{N_1}\sum_{i_2=1}^{N_2} \mu_{A_1^{i_1}}(x_1).\mu_{A_2^{i_2}}(x_2)}, \tag{2.18}$$

where $f(x_1, x_2)$ denotes the approximated output and $\{A_i^1, A_i^2, ..., A_i^{N_i}\}_{i=1,2}$ are fuzzy sets with $\{\mu_{A_1^{i_1}}(x_1)\}_{1 \le i_1 \le N_1}$ and $\{\mu_{A_1^{i_2}}(x_2)\}_{1 \le i_2 \le N_2}$ membership functions defined for inputs x_1 and x_2, respectively. The central desired outputs are denoted by $\bar{y}^{i_1 i_2}$. More information about the derivation steps of the fuzzy system is presented in [Wang 1994].

The output of the fuzzy system is passed through a gain control block that adaptively tunes the gain of the feedback loop according to the buffer size (or delay) and the video content properties as:

$$\Delta Q_F = G_F \times R_T / S_B \times f(x_1, x_2) \ , \tag{2.19}$$

where G_F is a coefficient which can be used for fine tuning of the RCA according to the video content properties. To tune G_F adaptively to the content, in the range, it can be roughly considered as a linear function of the average relative complexity as defined in (2.16).

Quality Controller

The quality controller computes an additive term to the final QP based on the quality of the previously encoded picture and a local average quality of encoded pictures. The idea is while the fuzzy controller provides the buffer constraint; the quality controller minimizes the variation in quality of the encoded video using the available buffer space. To compute the output of the quality controller ΔQ_Q, the average values (over encoded pictures in the current scene) of QP and PSNR are considered as a reference point. Then, it is attempted to drive the PSNR of subsequent pictures to the reference point with a gain proportional to the current deviation from the reference point. The output of the quality controller is computed by

$$\Delta Q_Q = \theta \times \bar{Q}(PSNR - \overline{PSNR}), \quad -1 \le \Delta Q \le 1 , \tag{2.20}$$

where $\bar{Q}$ is the average QP over the encoded frames in the current scene. $PSNR$ and $\overline{PSNR}$ are the PSNR of the previously encoded frame and the average PSNR over the encoded frames in the current scene, respectively. The PSNR values are computed based on the luminance component. The θ is a constant coefficient that defines the gain of the quality control loop. A larger gain for the quality control loop provides more constant quality while it may increase the fluctuations of buffer fullness. This means additional use of available space in the buffer. The value of ΔQ_Q is limited to the range of (-1, 1) to enforce the buffer constraint and to prevent instability in the control system.

In fact the fuzzy controller and the quality controller operate in parallel feedback loops that can have positive or negative outputs. If the outputs of the two controllers have different signs, it means there is a conflict between controlling the rate and controlling the quality. When the buffer conditions are critical, priority is given to the fuzzy rate controller to enforce the buffer constraint. Note that according to (2.20), the output of the quality controller is bounded in the range of [-1, 1] while the output of the fuzzy controller can vary over a wider range. Therefore, when the buffer conditions are critical, the output of the fuzzy controller is dominant and in the normal buffer conditions, depending on the quality measure, the quality controller can be dominant. Note also that it is possible to have a considerable increase or decrease in QP just by the quality controller because of integration of subsequent outputs after a number of frames.

Intra Prediction Pictures

The QP of I-picture is computed based on the picture complexity, target bit rate, buffer state, and scene cut information. Figure 7 depicts a block diagram for the calculation of QP for I-pictures. There are two types of I-pictures in the bit stream: periodic I-pictures which are placed in locations with a constant frequency and the I- pictures which are inserted at the beginning of scene cuts. The proposed QP for both types of I-pictures is formulated as:

$$Q_I = Q_R + A_C(\Delta Q_X + \Delta Q_B + \Delta Q_D), \tag{2.21}$$

where Q_I denotes the QP of I-picture. A_C is a constant value called content adaptation factor that can be used for fine tuning the rate control according to video content properties. Q_R is

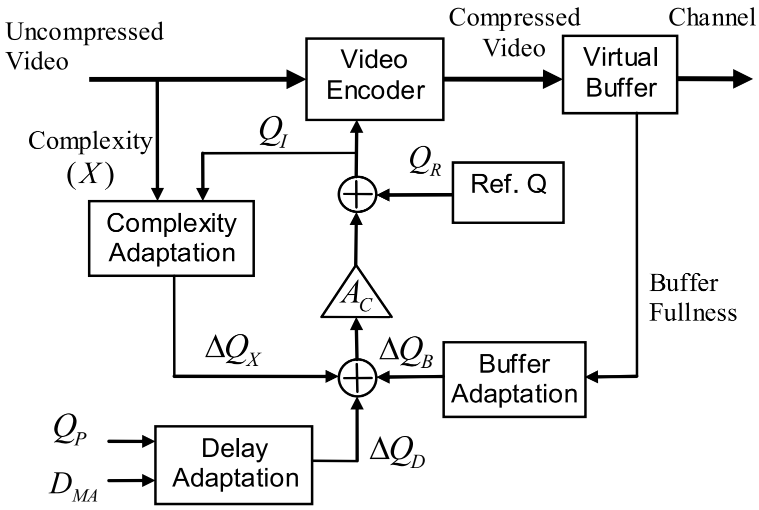

Figure 7: Block diagram of the RCA for the I-Pictures

a reference value for the Q_I computed differently for the two types of I-pictures. The control of Q_I or the variation of Q_I around the reference Q_R is imposed by three controlling signals ΔQ_X, ΔQ_B and ΔQ_D. The signals ΔQ_X and ΔQ_B adapt the QP of I-pictures according to the coding complexity of video picture and the occupancy of virtual buffer respectively. Moreover, ΔQ_D adapts the QP according to the target delay which is a function of target bit rate and the size of virtual buffer. While Q_R defines a reference value for the QP, the controlling signals make small variations around the reference value. To compute the small variations around the reference point, it is enough to use a simple first-order RD model. More details about the controlling signals and the reference QP are presented in [P5].

Simulation Results

The proposed semi-fuzzy RCA has been evaluated from different points of view using comprehensive simulations on various video contents encoded by different encoding parameters. The results of encoding have been compared with the constant QP encoding case and also with the JM RCA. According to the average simulation results, in comparison with the constant QP case, the semi-fuzzy algorithm has provided a lower delay, a higher PSNR (0.2dB), and a smaller average QP (0.2) at the expense of a larger standard deviation of PSNR. In comparison with the JM RCA, it provides a lower standard deviation for the PSNR, a smaller average QP (0.3) in a similar operating point in terms of PSNR and delay.

Results of more simulations show that the semi-fuzzy RCA can operate in a wide range of operating area between constant bit rate and constant quality. Moreover, graphical results show that there are strong correlations between the graphs of QP, PSNR and occupancy of buffer provided by the semi-fuzzy RCA algorithm. The strong correlation between the graphs means that the bit allocation is always implemented according to the complexity of the video content and available resources including the bit budget and delay.

From a computational complexity point of view, the proposed RCA has a remarkably low degree of complexity. Simulation results show that the proposed RCA has a computational complexity less than 4% of the JM RCA in terms of processing time. According to the provided results there is a big difference between the complexities of the two algorithms. The difference is even bigger when JM RCA is used with control at the MB level.

Chapter 3

Video Coding for Streaming over DVB-H

3.1. INTRODUCTION TO DVB-H

DVB-H (Digital Video Broadcasting for Handheld) is an ETSI specification for delivering broadcast services to battery-powered handheld receivers [ETSI EN302304], [Kornfeld 2004], [May 2004], [Faria 2006]. DVB-H is mainly based on the DVB-T specification for digital terrestrial television, adding to it a number of features to take into account the limited battery life of handheld receivers, and the particular mobile environments in which such devices must operate [ETSI EN300744], [Labedbusch 2006].

Comparing to digital terrestrial television, handheld television presents several technical challenges. First, mobile terminals have very small antenna size comparing to standard television antennas. Consequently, handheld terminals need stronger signals than standard television. Second, a handheld receiver is expected to work everywhere, especially inside buildings and in moving vehicles. This demands more robustness for transmission signal. Finally, handheld terminals use battery power supply with limited lifetime which restricts the working time and processing power. The DVB-H mobile TV standard addresses these issues by adding a number of features to DVB-T standard. DVB-H adds functional changes in the link and physical layers while it is backward compatible with DVB-T. In the link layer, DVB-H has two new elements when compared to DVB-T: *Time Slicing* and *Multiprotocol Encapsulation Forward Error Correction* (MPE-FEC). More details about these extensions are presented in the sequel. The extensions of DVB-H in the physical layer are not relevant to the subject of this thesis and hence not discussed here.

Time-Sliced Transmission

Services used in mobile handheld terminals require relatively low bit rates. The estimated maximum bit rate for streaming video services using advanced compression technologies is in the order of a few hundred Kb/s. A DVB transmission system usually provides a bandwidth of 10 Mb/s or more. This provides a possibility to significantly reduce the average power consumption of a DVB-H receiver by introducing a transmission scheme based on time division multiplexing. This scheme is called time-slicing. To reduce the power consumption in handheld terminals, the service data is time-sliced and then sent through the channel as bursts at a significantly higher bit rate compared to the bit rate of the audio-visual service. Time-slicing enables a receiver to stay active only a small fraction of the time, while receiving bursts of a requested service. This significantly reduces the power consumption used for radio reception parts. Figures 8 and 9 compare the continuous and the time-sliced transmission schemes in typical DVB-T and DVB-H channels. Figure 9 also shows the backward compatibility of DVB-H with DVB-T system in which a number of time-slice DVB-H services are multiplexed to one DVB-T transmission channel. Time-slicing also supports a quasi-optimum seamless handover by accomplishing the changing of the reception from one transport stream to another during the off-time between two bursts.

MPE-FEC

DVB-H employs an additional FEC to further improve mobile and indoor reception performance of DVB-T. In FEC a number of p extra symbols, called parity, are added to equal-length data symbols by a FEC code. In case of error, a FEC code can ideally reconstruct any p corrupted data symbols, when the location of the error is known and the number of corrupted symbols is less than $p/2$. The *Reed-Solomon* FEC code [Reed 1960] is a good example of an FEC code that is optionally used in DVB-H. Since Reed-Solomon code is a systematic code, that is, the source data remains unchanged after FEC encoding, MPE-FEC decoding is made optional for DVB-H receivers. MPE-FEC is computed over IP packets [Postel 1981].

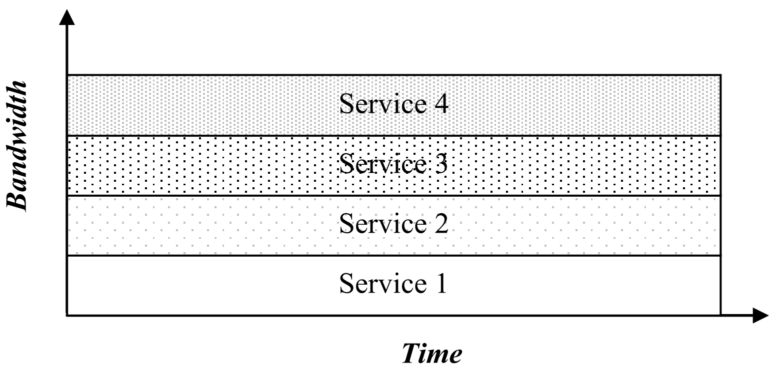

Figure 8. Continuous transmission of parallel services in DVB-T channel

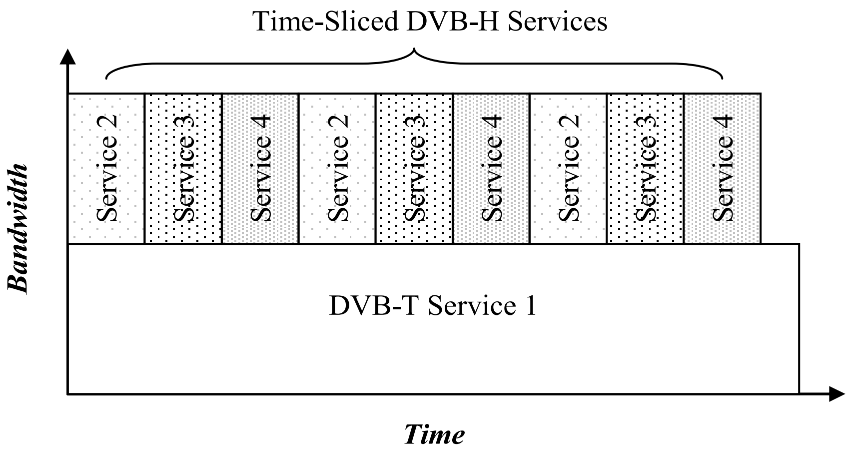

Figure 9. Time-sliced DVB-H services in parallel and compatible with DVB-T services

Time slicing and FEC are performed in a network element called *IP encapsulator*. The IP encapsulator encapsulates IP packets into *Multi-Protocol Encapsulation (MPE) Sections* which are further packetized into MPEG-2 *Transport Stream* packets. The optional FEC codes are also encapsulated into MPE-FEC sections. MPE sections are transmitted prior to MPE-FEC sections such that a receiver can just receive the unprotected data and ignore the protection data that follows.

Video Streaming over DVB-H

Video streaming over DVB-H channels significantly differs from other video streaming applications. The time-sliced transmission scheme is used in DVB-H to increase the battery life time of handheld receivers. The end-to-end delay of DVB-H services has increased due to the time-sliced transmission scheme. Depending on the time-slicing parameters, a lower end-to-end delay is possible by more power consumption at the receiver. Moreover, the end to end delay increases due to the variations that exist in the bit rate of the encoded media bit streams that are broadcasted. Variable bit rate video encoding is used to provide higher visual quality for broadcast services. For a given level of variations in bit rate, the end-to-end delay can decrease by a higher transmission bandwidth that means less utilization of bandwidth. On the other hand, the bandwidth utilization can be improved by statistical multiplexing of VBR broadcast services. Therefore, the video quality, end-to-end delay, bandwidth utilization, and power consumption of DVB-H receiver are related parameters. These parameters should be considered simultaneously when a DVB-H broadcast system is designed.

In this thesis, first the relationships between the end-to-end delay, the receiver power consumption, and the perceived video quality are explored in [P6] to provide some reference points for the system design and also for the video encoding and rate control task. According to the results provided in [P6], a delay constrained fuzzy RCA for DVB-H application is proposed [S3]. The proposed RCA is optimized for a DVB-H application to provide VBR bit streams with a relative constant visual quality and low end-to-end delay. In another major contribution of this thesis, a video encoding method for broadcast over DVB-H is proposed in [P7] that considerably decreases the end-to-end delay of DVB-H services. The contributions [P6] and [P7] are summarized in the sequel.

3.2. OPTIMAL CHANNEL SWITCHING DELAY IN IPDC OVER DVB-H

Channel switching delay or *tune-in time* in *IP data casting* (IPDC) over DVB-H refers to the time between the start of switching to a new service channel and the start of media rendering. Tune-in time is one of the significant factors in end-to-end delay of DVB-H broadcast system. Figure 10 shows a graphical demonstration for the channel switching delay in DVB-H when switching from Service 1 to Service 2. The shown delay times in the graph are: *Arrival Delay* (delay to arrival of desired burst), *Reception Delay* (reception duration of desired burst), *Decapsulation Buffering Delay*, *Decoder Refresh Delay* (delay to the first random access point, e.g. an *Instantaneous Decoder Refresh* (IDR) picture in H.264/AVC video coding standard or I-picture in other video coding standards), *Decoder Buffering Delay* (Initial buffering period of *Coded Picture Buffer*). The decapsulation buffering delay includes two buffering delays for the *Multi-protocol Decapsulation Buffer* (MDB) and RTP (Real Time Protocol) *Decapsulation Buffer* (RDB) [S4], [ETSI TS102472]. A buffering model for hypothetical DVB-H receiver has been proposed in [S4] that has been adopted in the standard specification [ETSI TS102472]. The decapsulation buffering delay is required to compensate the variations that exist in the burst size and the decoder buffering delay is needed to compensate the variations that exist in the bit rate of media bit streams. Moreover, another delay is needed for synchronization between the associated streams (e.g. audio and video) of the streaming session which is not shown in the graph.

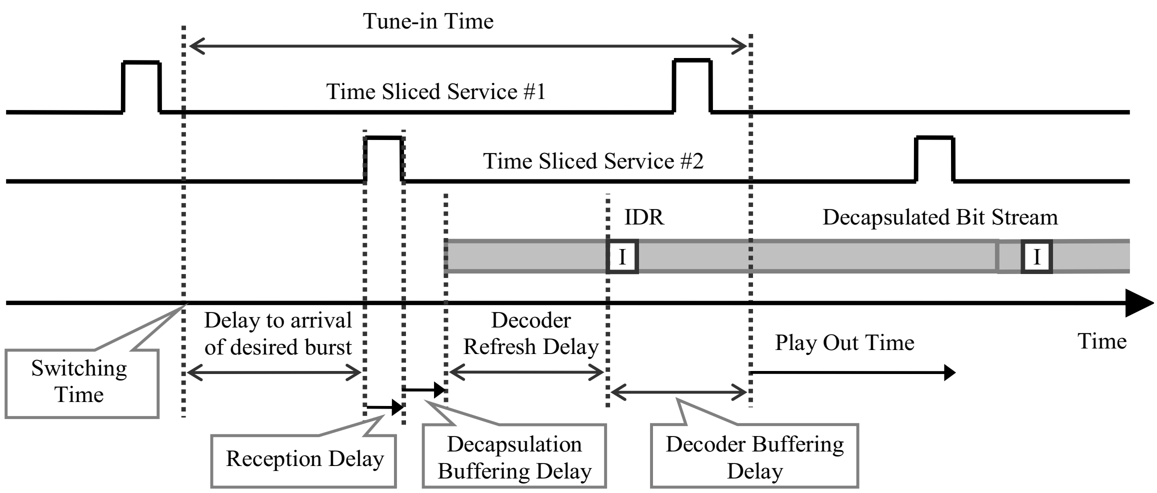

Figure 10: Channel switching timing graphs

Channel switching delay or tune-in time in IPDC over DVH-H is related to several essential parameters in the system including power consumption and video quality that are studied in [P6]. The results of the study are summarized in the sequel.

Power Consumption and Delay

In DVB-H, service data are time sliced and then sent to the channel as bursts at a significantly higher bit rate compared to the bit rate of the audio-visual service. This enables a receiver to stay active only a small fraction of the time, and thereafter decreases the consumed power for data reception. Figure 11 depicts the burst parameters in time-slicing scheme. *Burst Size* B_B refers to the number of network layer bits within a burst. *Burst Bit Rate* R_B is the bit rate used by a time-sliced elementary stream while transmitting a burst. *Constant Bit Rate* R_C is the average bit rate required by the elementary stream when not time-sliced. *Burst Duration* T_B is the time from the beginning to the end of the burst. *Off-Time* T_{off} is the time between subsequent bursts of the same service. The percentage of power saving resulting from time-slicing can be expressed as:

$$P_S = \frac{T_{off}}{T_{on} + T_{off}} \times 100 \tag{3.1}$$

where T_{on} denotes *On-Time* or the time duration when the radio receiver is on. For a given R_B and R_C, a closed form function for the percentage of power saving is derived in [P6] as:

$$P_S = \left(1 - R_C\left(\frac{1}{R_B(1-h)} + \frac{T_S}{B_B}\right)\right) \times 100 \tag{3.2}$$

where T_S denotes the *Synchronization Time* that is required by a receiver to re-acquire lock onto the signal before the start of the reception of the next burst. h is a constant value that compensates the overhead caused by the transport packet and section headers. Using (3.2) a closed form expression relating power saving and *receiving delay* (including the delay to arrival of the desired burst and the reception time of the burst) is given as:

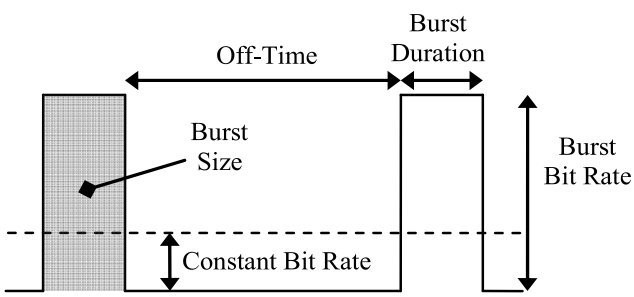

Figure 11: Burst parameters

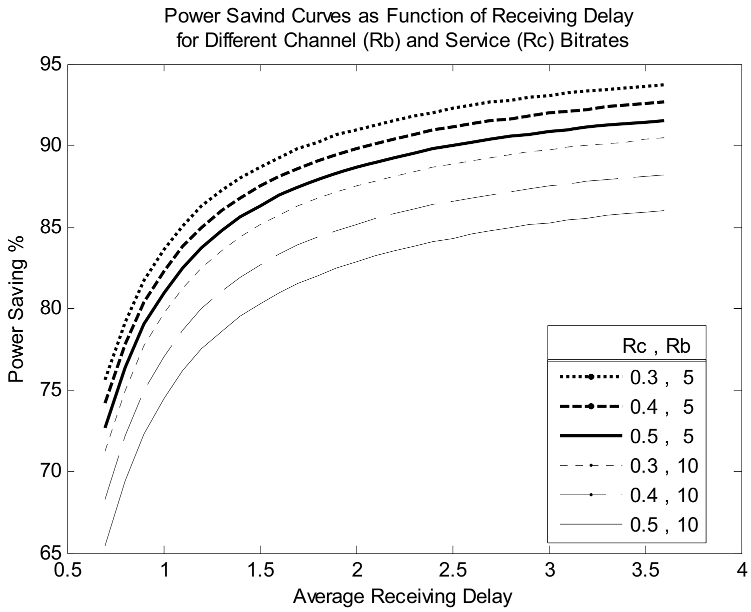

Figure 12: Power saving as a function of receiving delay

$$P_S = \left(1 - \frac{1}{D_R - Ts}\left(\frac{T_S}{2} + \frac{R_C D_R}{R_B(1-h)}\right)\right) \times 100\,. \tag{3.3}$$

were D_R denotes the receiving delay.

For transmission of an elementary stream with a desired average bit rate of R_C, by a transmission channel with bandwidth R_B, different operating points on the power saving curves (P_S, D_R) can be found from (3.3). Considering typical values for h (0.04 or 4%) and T_S (200 ms), a set of power saving curves are depicted in Figure 12. The power saving curves are computed for two different values of R_B (5 Mb/s and 10 Mb/s) and three different values of R_C (300, 400, 500 Kb/s). The results are used to find an optimal operating area in terms of receiving delay and the receiver power consumption.

Video Quality and Delay

In time-sliced services at least one random access point in each burst is desired. Therefore, the maximum value of the decoder refresh delay is related to the power saving percentage. The average value of decoder refresh delay, which decreases by frequent intra pictures in the bit stream, is constrained by the compression performance and video quality. For a given bit rate, when the number of intra pictures increases, the average quality of encoded bit stream degrades because, in comparison with inter prediction pictures, intra pictures need a higher bit budget for encoding in a similar quality. The decoder buffering delay is also related to the video quality. The buffering delay is defined according to variations that exist in the bit rate

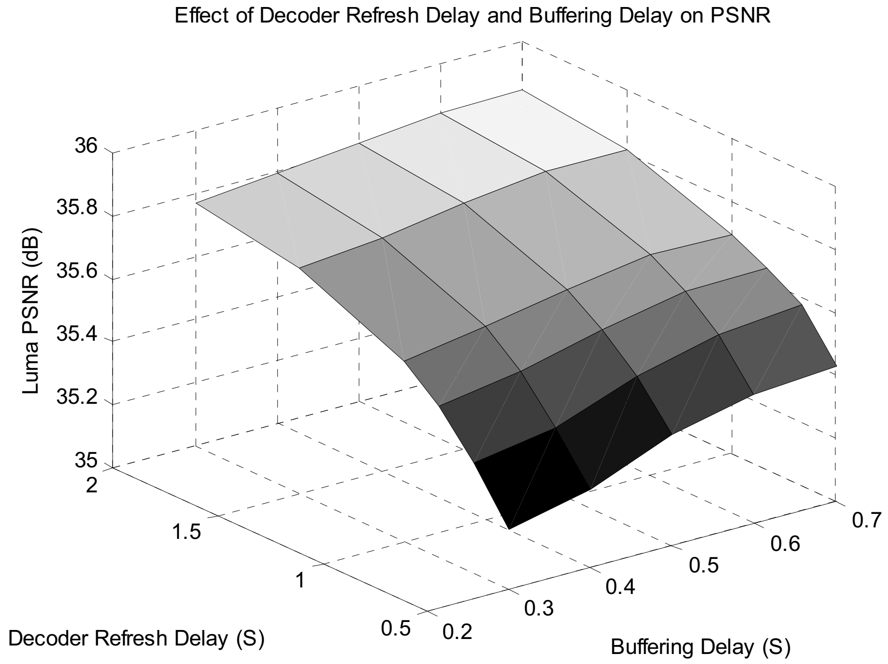

Figure 13: Effect of decoder refresh delay and buffering delay on PSNR

of the bit stream. A better average quality for compressed video is achieved at the expense of more variations in bit rate and thereafter higher buffering delay.

To find optimal values for the decoder refresh delay and buffering delay an empirical approach was used. A set of simulation was run in which a set of video sequences with various contents were encoded with different encoding parameters related to decoder refresh delay and buffering delay. For each encoded bit stream the PSNR of the luma component, the mean and the standard deviation of the QP of encoded frames were measured as quality criteria. Smaller values for the mean and the standard deviation of QP correspond to a higher quality. Sample simulation results are depicted in Figure 13 for a typical parameter set including: 300 kb/s bit rate, 15 fps frame rate, and QVGA picture format. The figure shows the average PSNR of compressed video, (luma components) as a function of two delays: decoder refresh delay and initial buffering delay.

According to simulation results an operating area is recommended in [P6] that provides optimal results in terms of tune-in time and video quality. The recommended values are provided based on simulation in which a single DVB-H service is transmitted through a CBR channel. The recommended values can be relaxed if statistical multiplexing is used for simultaneous transmission of a number of services.

3.3. TUNE-IN TIME REDUCTION IN VIDEO STREAMING OVER DVB-H

A simplified block diagram of a conventional IPDC system over DVB-H is depicted in Figure 14. As shown, a content encoder encodes an uncompressed source signal into a compressed media bit stream. Content encoder is typically capable of encoding more than one media type, e.g. audio and video. Alternatively, more than one content encoder may be required to code different media types of the source signal.

The coded media bit stream is transferred to a server. Examples of the format used in transmission include an elementary self-contained bit stream format, a packet stream format, or one or more coded media bit streams encapsulated into a container file. Content encoder and server may reside on the same physical device or may be included in separate devices. Content encoder and server may operate with live real-time content, in which case the coded media bit stream may not be stored permanently, but rather buffered for small periods of time in content encoder and/or in server to smooth out variations in processing delay, transfer delay, and coded media bit rate. Content encoder may also operate off-line considerably earlier than when the bit stream is transmitted from the server. In such a case, the system may include a content database, which may reside on a separate device or on the same device as the content encoder and/or server.

The server may be an IP multicast server using real-time media transport over RTP. The server is configured to encapsulate the coded media bit stream into RTP packets according to an RTP payload format. Although not shown in this figure, the system may contain more than one server. The server is connected to an IP encapsulator. The connection between the server and the IP encapsulator may be through the IP network or a fixed-line private network. The IP encapsulator encapsulates IP packets into MPE Sections which are further packetized into MPEG-2 Transport Stream packets. The IP encapsulator optionally uses MPE-FEC based on Reed-Solomon codes. An IPDC system over DVB-H further includes at least one radio transmitter which is not discussed here. To reduce power consumption in handheld terminals, the service data is time-sliced and then sent to the channel as bursts. Finally, the system includes recipients, capable of receiving, demodulating, decapsulating, decoding, and rendering the transmitted signal, resulting into uncompressed media data.

One of the critical factors in tune-in time is the time until a media decoder is refreshed to produce correct output frames, which can be minimized if an MPE-FEC frame is started with

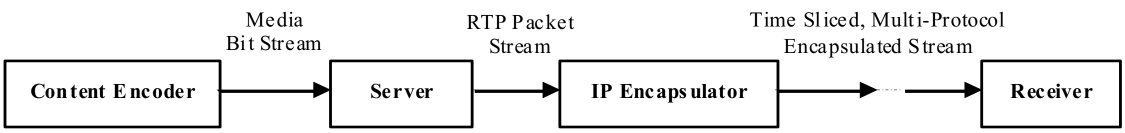

Figure 14: Simplified block diagram of a conventional IPDC system

a random access point, such as an IDR picture in H.264/AVC. It should be noted that if the decoder started decoding from an IDR picture that is not at the beginning of a time-slice immediately when the time-slice is received, the input buffer for decoding would drain before the arrival of the next time-slice and there would be a gap in video playback corresponding to the play out duration from the beginning of the time-slice to the first IDR picture.

In IPDC over DVB-H, the content encoding and the encapsulation to MPE-FEC frames are implemented independently and it is hard to control the exact location of IDR pictures relative to the boundaries of MPE-FEC frames. Using very frequent IDR pictures in the bit stream may reduce the average decoder refresh delay. However, frequent IDR pictures in the bit stream drop the compression efficiency remarkably.

A method for fast channel zapping in *Set-Top Box* applications has been presented in [Boyce 2005] in which an auxiliary bit stream including frequent low quality IDR pictures is sent to the receiver in parallel to the main bit stream. When channel change the receiver replaces an inter picture from the main bit stream with an IDR picture from the auxiliary bit stream. Although this method can decrease the average tune-in time in IPCD over DVB-H, it cannot minimize the tune-in time, as the IDR picture intervals in the auxiliary bit stream may not match with time-slice intervals. Furthermore, the auxiliary bit stream consumes a considerable amount of transmission bandwidth. Moreover, some modifications in the receiver are required relative to the DVB-H specifications to switch between two bit streams. Finally, the employed low quality IDR pictures degrade the quality of the subsequent pictures up to the next normal IDR picture.

In [S5], we proposed a video splicing method, which minimizes the decoder refresh time in IPDC over DVB-H without any increase in bandwidth and no modifications at the receiver, at the expense of a reasonable degradation in video quality. We proposed a modification in the operation of IPDC system, which minimizes the decoder refresh delay by an IDR picture splicing method. The proposed method is implemented partially at the encoder and at the IP encapsulator. According to the proposed video splicing method, an additional stream consisting of refresh pictures only is encoded and transmitted to the IP encapsulator. The IP encapsulator replaces some pictures in a normal bit stream with the refresh pictures according to time-slice boundaries in order to achieve the minimum decoder refresh delay. Replacing pictures causes a mismatch in the pixel values of the reference pictures between the encoder and decoder and the mismatch error is propagated in the reconstructed video. It has to be ensured that the propagated error is subjectively negligible. Furthermore, the “spliced” stream resulting from the operation of the IP encapsulator should comply with the HRD specification of H.264/AVC. In [S5], the propagated error was investigated to verify that it is subjectively negligible. Moreover, a video rate control system is proposed in [S6] to guarantee the standard compliancy of the spliced bit stream. The proposed rate control system is implemented partially in both the content encoder and IP encapsulator. Details related to the rate control system at the content encoder was presented in [S6] while those related to the rate control system at the IP encapsulator were presented in

[S7]. Furthermore, some details about the buffering requirements of spliced video were presented in [S8]. An overview on the proposed video encoding and splicing method and more investigation results are presented in [P7]. The proposed video encoding and splicing method is summarized in the sequel.

3.3.1. Video Encoding and Splicing

A simplified block diagram of the proposed IPDC system is depicted in Figure 15. At the content encoding level, two video encoders encode the uncompressed video to two encoded primary bit streams including a *Spliceable Bit Stream* (SBS) and a *Decoder Refresh Bit Stream* (DRBS) from the same source picture sequence. The SBS includes frequent spliceable pictures which are reference pictures constrained as follows: no picture prior to a spliceable picture is referred to in the inter prediction process of any reference picture at or after the spliceable picture, in decoding order. Non-reference pictures after the spliceable picture may refer to pictures earlier to the spliceable picture in decoding order. These non-reference pictures cannot be correctly decoded if the decoding process starts from the spliceable picture, but can be safely omitted as they are not used as reference for any other pictures. The DRBS contains only intra or IDR pictures corresponding to spliceable pictures and with a picture quality similar to the corresponding pictures. The DRBS and the SBS are transmitted from the server to the IP encapsulator. The IP encapsulator composes MPE-FEC frames, in which the first picture in decoding order is an IDR picture from the DRBS and the other pictures are from the SBS. The IDR pictures at the beginning of MPE-FEC frames minimize the tune-in time for newly-joined recipients. No changes in the receiver operation are required in the proposed system.

Replacing an inter picture with an intra picture in the SBS causes a mismatch in the pixel values of the reference pictures between the encoder and decoder. The mismatch propagates until the next IDR picture in the spliced stream. A technically elegant solution to avoid mismatch altogether would be to use SP (*Switching P*) and SI (*Switching I*) pictures of H.264/AVC, but they are only included in the extended profile of H.264/AVC [Karczewicz 2003], [Setton 2005]. The extended profile of H.264/AVC is not allowed in the current DVB-H standard [ETSI TS102005]. A similar solution was proposed for stream switching in streaming servers by Farber and Girod in [Farber 1997]. They used a type of S picture to enable switching between bit streams. Unlike SP pictures in [Karczewicz 2003] , S pictures introduce mismatch error while switching between bit streams. To minimize the mismatch error, the QP used for S pictures should be kept small. While in our application we do not need to switch between multiple bit streams, instead of high quality switching frames, the spliceable pictures and corresponding IDR pictures are proposed to be encoded with a higher quality than other pictures to decrease the mismatch error. This alternative solution for the DRBS has been studied and it is shown that it reduces the mismatch error remarkably [S5], [P7].

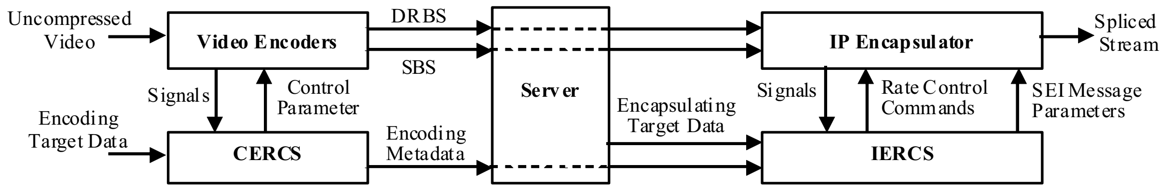

Figure 15: Block diagram of proposed splicing and rate control method

In addition to the mismatch of pixel values, the proposed method involves two buffering related challenges. First, the standard HRD compliancy of the spliced stream is hard to verify in the IP encapsulator. Second, the initial delay for coded picture buffering in the HRD is hard to derive in the IP encapsulator. More details about these technical challenges are discussed next.

3.3.2. Mismatch Error

To study the propagation of mismatch error in spliced bit stream, a comprehensive set of simulation was run on different video sequences with various encoding and splicing parameters while the propagated error was measured by several criteria such as PSNR, maximum sample-wise absolute difference, *Bjontegaard Delta PSNR* and *Bjontegaard Delta Rate* [Bjontegaard 2001].

As an important result, the degradation in quality of spliced streams in comparison to the spliceable bit streams saturates to a constant value after a small number of frames. This means that the degradation in quality is independent of IDR frame frequency in the spliced bit stream. Hence, using just one IDR picture in each MPE-FEC frame independently of MPE-FEC frame size can minimize the tune-in time and also the average bit rate. In other words, it is possible to have very large MPE-FEC frames each including only one intra/IDR picture with minimum tune-in time at the expense of a constant degradation in the quality of decoded video [S5][P7].

Simulation results show that when all pictures in SBS and DRBS are encoded with a similar QP, the spliced bit stream has a PSNR of about 1.6 dB less than SBS on the average over different video contents and different bit rates. When spliceable pictures and corresponding IDR pictures are encoded with a higher quality than other pictures, the average degradation in quality is reduced to 1.0 dB. Despite of the drop in PSNR, the results of a small-scale visual quality test indicated that non-expert viewers do not typically perceive the error.

Coding frequent and periodical IDR pictures in the transmitted bit stream is an alternative means to reduce the tune-in time. However, a penalty, much higher in term of quality degradation, should be paid for a similar tune-in time reduction. Typical results show that if the IDR period is changed from 30 to 5 frames to decrease the average decoder refresh delay from 1.0 to 0.17 seconds, respectively, the quality of encoded video degrades about 4.7 dB in average PSNR. Considering the simplicity of the proposed splicing method, and the penalty above, the proposed splicing method can be utilized when the use of SP and SI pictures of H.264/AVC is disallowed.

3.3.3. Video Encoding and Rate Control Method

According to the proposed splicing method the spliceable pictures and corresponding intra/IDR pictures in two primary streams should be encoded with similar qualities. In a similar quality, an IDR frame can consume a bit budget from 5 to 10 times more than the corresponding inter picture. Furthermore, similar quality for corresponding frames in two primary streams means that only the bit rate of one primary stream can be controlled. Consequently, there is no real short-term control on the bit rate of spliced streams, and therefore it is hard to verify the HRD compliancy of spliced streams. Moreover, the encoding parameters cannot be controlled according to the results of splicing, because encoding and splicing are performed independently and without any feedback link. Results of HRD simulation on a number of spliced bit streams confirm that the problem of HRD compliancy of spliced bit stream is serious and it is impossible to control the bit rate of spliced bit stream just by dropping frames in IP encapsulator [S8].

To solve the problem above, a comprehensive rate control system as depicted in Figure 15 is proposed which is implemented in both the content encoder and the IP encapsulator. The content encoding rate control system (CERCS) controls the bit rate of two primary streams considering a fixed value for the frequency of IDR pictures in a desired spliced stream. However, the frequency of IDR pictures in the spliced stream can have variations around a target value since the number of video pictures in MPE-FEC frames is not fixed. Moreover, in offline encoding the IDR picture frequency which has been used for the rate control of primary streams at the content encoder may be different from the average IDR frequency of spliced stream. The IP encapsulating level rate control system (IERCS) implements another control to compensate for the above variations and to provide HRD compliancy for the spliced bit stream. Furthermore, the H.264/AVC *Supplemental Enhancement Information* (SEI) message parameters related to buffering of the spliced bit stream can be provided by IERCS.

The CERCS controls the bit rate of primary streams according to the encoding target data which is set by the user and also according to several signals which are extracted from the uncompressed and compressed video. The encoding target data include target bit rate of spliced stream and average frequency of IDR pictures in the desired spliced stream. Furthermore, some encoding metadata as complementary information are provided by CERCS which are sent to the server and then to the IP encapsulator.

The IERCS controls the bit rate of the spliced stream according to the encoding metadata and encapsulating target data defined by the server. The encapsulating target data includes target bit rate of the spliced stream and IDR picture frequency of the spliced stream. Details of the proposed rate control systems are presented in the sequel.

Content Encoder Rate Control System

Figure 16 illustrates the block diagram of the proposed CERCS. The CERCS is configured to control the bit rate of the spliced stream by controlling the bit rate of SBS taking into consideration the bit rate of DRBS. In fact, the CERCS tries to simulate the spliced bit stream that is built at the IP encapsulator later on and then, it controls the bit rate of SBS and DRBS according to the bit rate of the simulated spliced bit stream. Moreover, the CERCS minimizes the changes of encoding parameters to provide high visual quality for encoded video. According to the diagram, two separate video encoders, Encoder 1 and Encoder 2, encode the uncompressed video to provide SBS and DRBS, respectively. Two virtual buffers, Virtual Buffer 1 and Virtual Buffer 2, are utilized in this system. Virtual Buffer 1 provides buffering constraints for the target spliced stream and Virtual Buffer 2 moves the extra bit rate resulting from replacing spliceable pictures by IDR pictures to Virtual Buffer 1 gradually to minimize short-term fluctuations in encoding parameters and to maximize the quality of encoded video. The VBR video rate controller shown in the diagram can be any RCA with buffer constraint, such as the proposed algorithms in [P1], [P5], and [S2]. This is an advantage for the system that it can utilize different available video RCAs without any modification. The QP computed by the rate controller is used by the two encoders. The target bit rate for the rate controller which is fixed in normal applications, is utilized as a control signal in this system. The target bit rate is provided by the TBRE (target bit rate estimator) block. The TBRE adjusts the target bit rate of the spliceable bit stream, as a control signal smoothly, according to the target bit rate of the spliced stream and the results of encoding DRBS. More details about the virtual buffers and TBRE are presented in [P7].

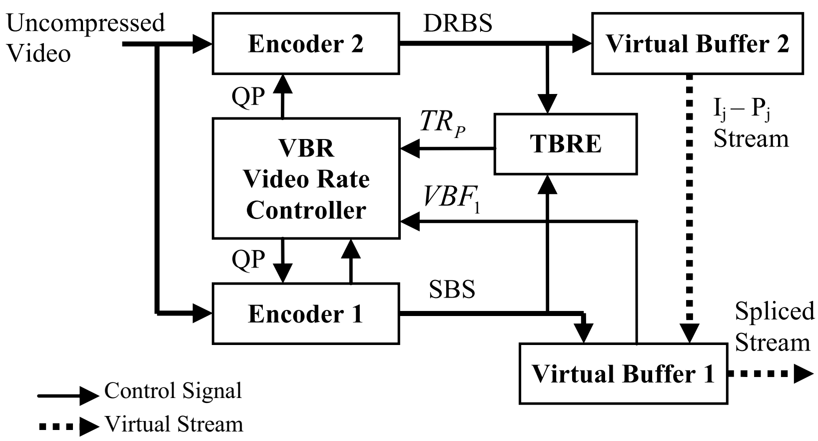

Figure 16: Block diagram of rate control system at content encoder level (CERCS).

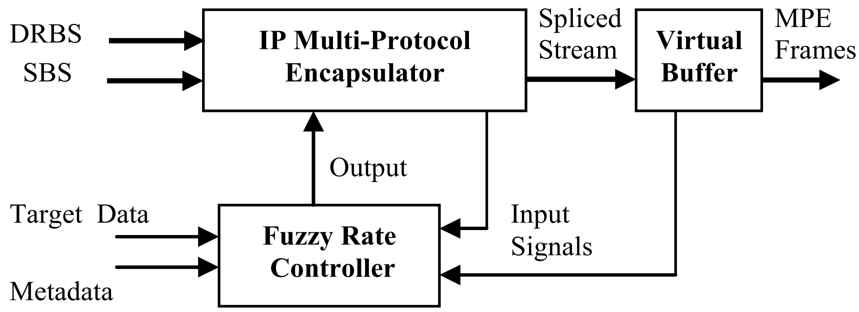

Figure 17: Block diagram of proposed IERCS

IP Encapsulator Rate Control System

Figure 17 illustrates the block diagram of the proposed IERCS. The IERCS utilizes a fuzzy rate controller and a Virtual Buffer. The fuzzy controller controls the bit rate of the spliced bit stream by controlling the frame rate and the type of pictures. It may drop a number of pictures from an MPE-FEC frame to decrease the bit rate or it may replace one or more extra spliceable pictures by IDR pictures to increase the bit rate. The fuzzy controller and virtual buffer operate based on MPE frame. The size of the virtual buffer is computed according to the metadata and encapsulating target data.

The fuzzy controller is configured to provide control over the bit rate of the spliced stream while it minimizes the number of dropped pictures and prevents unnecessary IDR pictures. The output of the controller is a natural number. A positive number shows the number of pictures that should be dropped from the start or end of MPE frame and a negative number shows the number of extra spliceable pictures which can be replaced by IDR pictures in MPE frame. Locations of extra IDR pictures are distributed uniformly along the MPE frame. The fuzzy controller uses the following two signals as inputs:

$$INPUT1 = FB \times FR / TR_S - F \,, \tag{3.4}$$

$$INPUT2 = (BF - I_j + P_j) / BS \,, \tag{3.5}$$

where FB denotes the total number of bits consumed by the current MPE frame before any dropping or extra IDR picture. BF and BS refer to the fullness and size of the virtual buffer, respectively. TR_S and FR are the target bit rate and frame rate of the spliced stream, respectively. F denotes the target number of pictures in one MPE frame in the target frame rate. I_j and P_j stand for the bit budgets consumed by the j^{th} replaced IDR picture and the corresponding spliceable picture, respectively.

All fuzzy rules are summarized in Table 3. The content of the table specifies the output of the controller. The letters H, L, M, V, X and S correspond to the fuzzy descriptions High, Low, Medium, Very, Extremely, and Super, respectively. The desired central values for the output

TABLE 3. SUMMARIZATION OF IF-THEN FUZZY RULES

INPUT1 \ INPUT2	VL	L	ML	M	MH	H	VH
XL	M	M	M	M	ML	L	VL
VL	M	M	M	M	M	ML	L
L	M	M	M	M	M	M	ML
ML	M	M	M	M	M	M	M
M	MH	M	M	M	M	M	M
MH	H	MH	M	M	M	M	M
H	VH	H	MH	M	M	M	M
VH	XH	VH	H	MH	M	M	M
XH	SH	XH	VH	H	MH	M	M

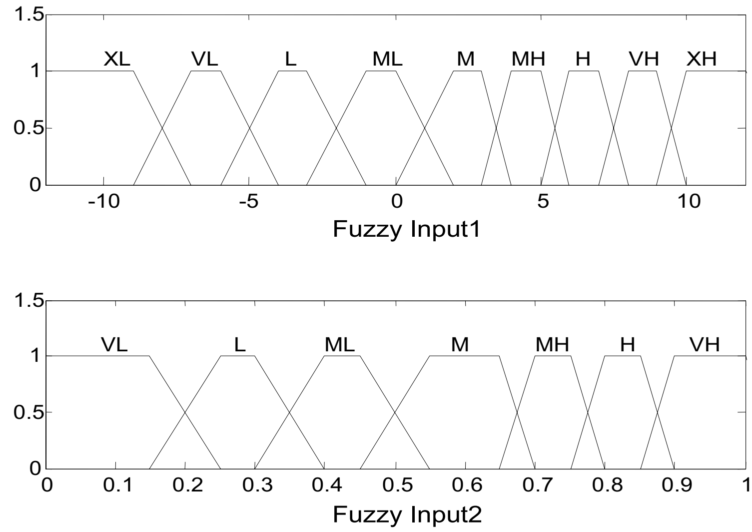

Figure 18: Membership function of the linguistic variables

of the fuzzy system correspond to VL, L, ML, M, MH, H, VH, XH, SH in the table are -3, -2, -1, 0, 1, 2, 3, 4, and 5, respectively. The shape of fuzzy membership functions are shown in Figure 18. A fuzzy system with two inputs using product inference engine, singleton fuzzifier, and centre average defuzzifier as (2.18) was used.

Simulation Results

Typical intervals between time-slices containing content for a particular audio-visual service may range from one second to a couple of seconds. If IDR pictures are placed randomly in a normal bit stream and the average IDR picture interval is equal to the time-slice interval, the expected tune-in time due to a decoder refresh is approximately half of the time-slice interval, i.e. typically from half a second to few seconds. From tune-in time reduction point of view, the proposed splicing method can typically decrease the decoder refresh time to very close to zero or even to zero. This reduction of tune-in time is obtained at the expense of a relatively small degradation in PSNR of the spliced video resulting from the mismatch error as discussed already.

To evaluate the performance of the proposed CERCS, the quality of the spliceable video by this system were compared with the quality of similar bit streams which are encoded by an independent encoder and RCA. Simulation results on various video contents show that corresponding encoded bit streams by the two systems have exactly similar average qualities in term of PSNR.

The performance of the proposed rate control system was also evaluated from the standard compliancy point of view and also from the picture drop rate point of view. Simulation results show that the proposed system can provide standard compliant bit streams while the picture drop rate is very small (less than 0.3%).

The overall simulation results show that the proposed splicing method and the rate control system can provide standard compliant video bit streams for streaming over DVB-H with good average quality and with minimum tune-in time and complexity at the expense of a relatively small degradation in quality of the spliced video.

Chapter 4

Joint Video Coding and Statistical Multiplexing

In video broadcasting, when a number of services are encoded and broadcasted simultaneously and the encoders are collocated geographically, it is possible to use a joint rate controller for all encoders instead of using independent rate controllers, one for each encoder. A joint rate controller can improve the average quality of encoded bit streams, and it can also decrease the end-to-end delay of a broadcast system by sharing resources among the bit streams. In fact, the joint rate controller imposes a type of statistical multiplexing (StatMux) on the encoded bit streams. In StatMux, a fixed bandwidth communication channel is virtually divided into several variable bandwidth channels. The resulting channels are adapted to the instantaneous traffic demands of the bit streams that are transferred over the channels. StatMux is used in many communication applications to improve the overall performance of communication channels. The performance of a communication channel can be evaluated in terms of bandwidth utilization, transmission delay, and data drop rate.

In joint video encoding, a common bit budget is shared between the bit streams according to their temporal complexities by a joint rate controller. The major problem of joint video encoding and StatMux is how to allocate the available bit budget among the video sources that share the common channel bandwidth. Several joint video encoding methods have been presented in the literature. The presented methods follow two main approaches. In the first approach, a preprocessing is performed on video sources to gather statistics about the coding complexity. Then, the real coding process can operate based on the statistics obtained by the pre-analysis. In the second approach that is based on a RD model, first it is attempted to model the performance of the video encoder and the coding complexity of video sources and then the allocated bits to video sources is controlled based on the resulting models while the models are updated during encoding. See the proposed methods in [Boroczky 2000], [Wang 1999] and [Xiong 2004] as examples for the two approaches. The system presented in [Boroczky 2000] consists of several preprocessors and video encoders. Each preprocessor analyzes a video source and derives picture statistics. Using these statistics, a joint rate controller calculates the bit rate for each encoder based on the relative complexities of the sources. Another method for joint video coding of multiple video sources is presented in [Wang 1999]. In this method, the input video sources are divided into *Super GOPs* (a number of GOPs, one from each source) and *Super Frames* (a collection of frames, one from each source) and then, the bit budget is distributed hierarchically between the video super GOPs, super frames and frames according to their relative complexities. Finally, using a RD model, a QP is calculated for each frame according to the allocated bits to the frame. A similar approach to that in [Wang 1999] is presented in [Xiong 2004].

In this thesis, two new methods for joint video encoding and StatMux are proposed in [P9] and [P8]. According to the proposed methods, the bandwidth can be shared for a long-term such that the average bit rate of bit streams can be changed in the long term or the bandwidth can be shared only for a short-term, such that the long term average bit rate of bit streams are constant. In comparison with independent encoding, in a similar buffering delay, the joint encoding with a long-term bandwidth sharing can provide a more constant quality for encoded bit streams. On the other hand, in comparison with independent encoding, in a similar quality, joint encoding with a short-term bandwidth sharing can provide a lower buffering delay for the bit streams. The proposed methods are tested for DVB-H application. However, they are applicable for other video broadcasting applications. The proposed joint video encoding methods in [P8], [P9] are summarized in the sequel.

4.1. JOINT VIDEO ENCODING WITH A COMMON QUALITY OF SERVICE

A novel fuzzy joint rate controller is proposed in [P8] that can be used for implementation of StatMux in conjunction with video encoding to improve the average quality of broadcast services. Furthermore, it can be used to decrease the end-to-end delay of broadcast system. The proposed method is different from the previous studied methods from several points of view. First, it is a real-time method without any look ahead even for one frame. Second, unlike the previous methods, it does not use any bit allocation strategy. The bit rates of bit streams are controlled by QP on a picture basis while the controller does not compute the QP directly. Utilizing a fuzzy controller, the controller only computes the variations of QP. Furthermore, thanks to the fuzzy controller, the proposed method has a very low degree of complexity in comparison to the studied methods. Moreover, it has a granular structure such that it can be used for encoding any number of bit streams. Finally, it is tuned easily to achieve the desired bit rate, quality, and delay. The joint rate control algorithm provides a near constant (along the bit streams) and common (across the bit streams) quality for the encoded video bit streams. Furthermore, it controls the variations in the bit rate of the aggregated bit stream to decrease the end-to-end delay of the broadcast system.

A simple block diagram of the proposed video encoding method is depicted in Figure 19. As shown, a number of video sources are encoded simultaneously to bit streams by separate encoders. A joint video rate controller (JVRC) controls the bit rate and the quality of encoded bit streams by computing a QP on a picture basis for each encoder. The JVRC utilizes a virtual joint buffer and a multiplex simulator (MUX SIM). The multiplex simulator simulates the ideal case of StatMux which is implemented later on by the IP encapsulator in DVB-H application. With this structure in which no feedback from multiplexer is used, the bit streams can be encoded independently of multiplexing. The joint buffer is charged by the multiplexed bit streams and drained by a constant bit rate equal to the channel bandwidth. However, it is possible to consider that the virtual buffer is running on the receiver side. In that case, it is charged by a constant bit rate equal to bandwidth and discharged by the multiplexed bit streams. The JVRC uses a feedback signal from the occupancy of joint buffer and a feedback from the bit rate of the aggregated bit stream to control the overall bit rate. It also uses other feedback signals from the distortion (PSNR) of encoded bit streams to distribute the available bandwidth between the bit streams proportional to their coding complexity. The JVRC can be functionally divided into two separate controllers: a *Fuzzy Rate Controller* and a *Quality Controller*. An internal block diagram for JVRC is depicted in Figure 20. More details about the joint encoding system are presented below.

Virtual Joint Buffer

Considering the joint buffer at the receiver side, the occupancy of the joint buffer is updated after encoding a series of corresponding pictures, one picture from each video source, called a *Super Picture* as:

$$O(m) = O(m-1) - \sum_{i=1}^{N} b_i^m + \frac{R}{F}, \tag{4.1}$$

where O denotes the occupancy of the joint buffer and m shows the index of the super picture. F stands for the frame rate and R indicates the target bit rate for the aggregated bit stream.

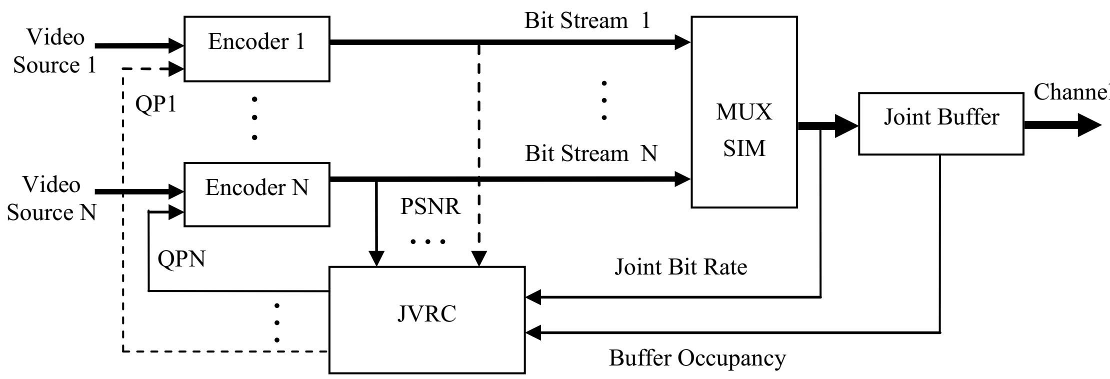

Figure 19: Block diagram of proposed video encoding method

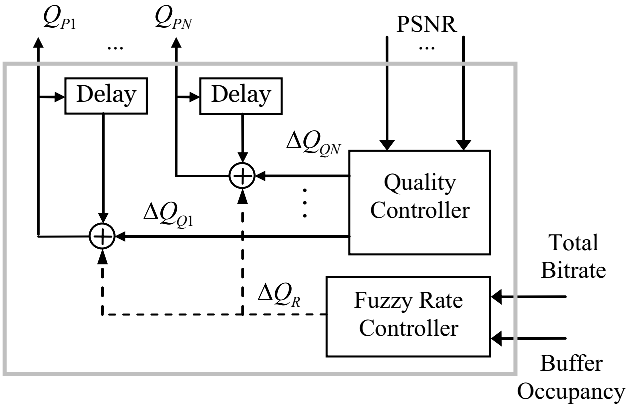

Figure 20: Internal diagram of JVRC

Joint Video Rate Controller

The JVRC computes a QP for each encoder based on the output of fuzzy controller (ΔQ_R), the corresponding output of the quality controller (ΔQ_{Qn}) and the QP is used for encoding the previous picture in the encoder. The QP for each encoder is computed as:

$$Q_n(i) = Q_n(i-1) + \left[\Delta Q_R + \Delta Q_{Qn}\right], \tag{4.2}$$

where n and i denote the indexes of the encoder and the picture, respectively. $[y]$ stands for the nearest integer to y. The outputs of the fuzzy controller and the quality controller are small values around zero which are added to the QP used for encoding the previous picture. In this structure, only the variations of QP are computed. The Fuzzy controller has been optimized to minimize the variations of QP and thereafter to provide a more constant quality for the bit streams. The quality controller tries to balance the quality across the bit stream.

Fuzzy Rate Controller

The fuzzy controller has been designed in such a way as to suppress the fluctuation of QP as much as possible while the buffer constraint is observed. The fuzzy controller has two input signals:

$$x_1 = O/S, \tag{4.3}$$

$$x_2 = \frac{I_I + X_{IP} - 1}{I_I} \cdot \frac{F}{R}\left(\sum_{i=1}^{N_P} b_{Pi} + \frac{1}{X_{IP}}\sum_{i=1}^{N_I} b_{Ii}\right), \tag{4.4}$$

where x_1 and x_2 correspond to the inputs of the fuzzy system. O and S stand for the occupancy and the size of the joint buffer, respectively. b_{Pi} and b_{Ii} denote the consumed bit budgets by the i^{th} P and I-picture, respectively, in the last encoded super picture. N_P and

N_I are the number of P and I-pictures in the last encoded super picture, respectively, while i shows the index of picture in the super picture. I_I indicates the interval of periodic I-pictures in the bit streams. F stands for the target frame rate. X_{IP} shows the average relative complexity of I-pictures to P-pictures and it is defined as:

$$X_{IP} = \overline{b}_I / \overline{b}_P, \tag{4.5}$$

where $\overline{b}_P$ and $\overline{b}_I$ denote the consumed bit budgets by the encoded P and I-pictures respectively in average over all encoded pictures. The input x_1 is a representation of buffer conditions. When the buffer is empty, $x_1 = 0$ and when the buffer is full, $x_1 = 1$. The input x_2 is the ratio of the instantaneous bit rate to the target bit rate. The instantaneous bit rate is normalized according to the relative complexity and the combination of P and I-pictures in the super pictures to ignore the variations in the bit rate resulting from high quality I-pictures.

A fuzzy system exactly similar to the one used in the semi-fuzzy RCA was used without any further optimization. Although the inputs to the system here differ from the semi-fuzzy RCA, the same fuzzy rules and MSFs were used.

Quality Controller

The quality controller computes an output for each encoder according to the quality of encoded bit streams to balance the quality across the bit streams. To compute the outputs of the quality controller, the proposed quality controller in the Semi-Fuzzy RCA has been modified for multiple bit streams. The output of the quality controller used in (4.2) is proposed to be computed by:

$$\Delta Q_{Qn} = \theta \cdot \bar{Q}\left(PSNR_n - \overline{PSNR}\right), \tag{4.6}$$

where $\bar{Q}$ denotes the average of QP over all pictures in the last encoded super picture. $PSNR_n$ stands for the PSNR of the last encoded picture in the n^{th} bit stream and $\overline{PSNR}$ is the average of PSNR over all pictures in the super picture. θ is a constant coefficient which is defined experimentally. To suppress the fluctuations of QP resulting from short-term variations in PSNR, a simple low pass filter such as (2.17) is used to smooth the variations in $\bar{Q}$ and $\overline{PSNR}$ before using them in (4.6).

Simulation Results

The performance of the proposed joint video coding and StatMux method is evaluated through simulations over 6 video sequences with typical encoding parameters corresponding to a DVB-H application. From video quality point on view, the proposed method has been compared with independent video rate control and also with the constant QP encoding.

Simulation results show that the proposed method can provide a more constant quality than independent rate control along the bit streams and a more constant quality than constant QP encoding across the bit streams. The graphs depicted in Figure 21 show the PSNR of encoded bit streams in the three cases. The graphs related to each bit stream are depicted with a different color. The graphs show how well the qualities of joint encoded bit streams are driven close together while they have only smooth variations comparable to the constant QP case.

The performance of the proposed joint encoding method has been compared with independent encoding from end-to-end delay point of view. According to simulation results, the joint encoding method can decrease end-to-end delay in a broadcast system. However, the overall performance depends on the size of the joint buffer and also on the number of joint encoded bit streams.

From computational complexity point of view, the proposed joint controller is comparable with the semi-fuzzy RCA that has a complexity less than 4% of JM RCA [P5]. The computational complexities of conventional joint RCAs increase proportionally to the number of joint encoded bit streams, but in the proposed method the main part of computational complexity that is related to the fuzzy controller is independent of the number of bit streams.

The overall results show that the proposed joint encoding method can decrease end-to-end delay of a broadcast system and improve the overall quality of broadcast services. It has a much lower computational complexity compared to existing techniques. Moreover, it has a granular structure such that it can be applied to any number of video services.

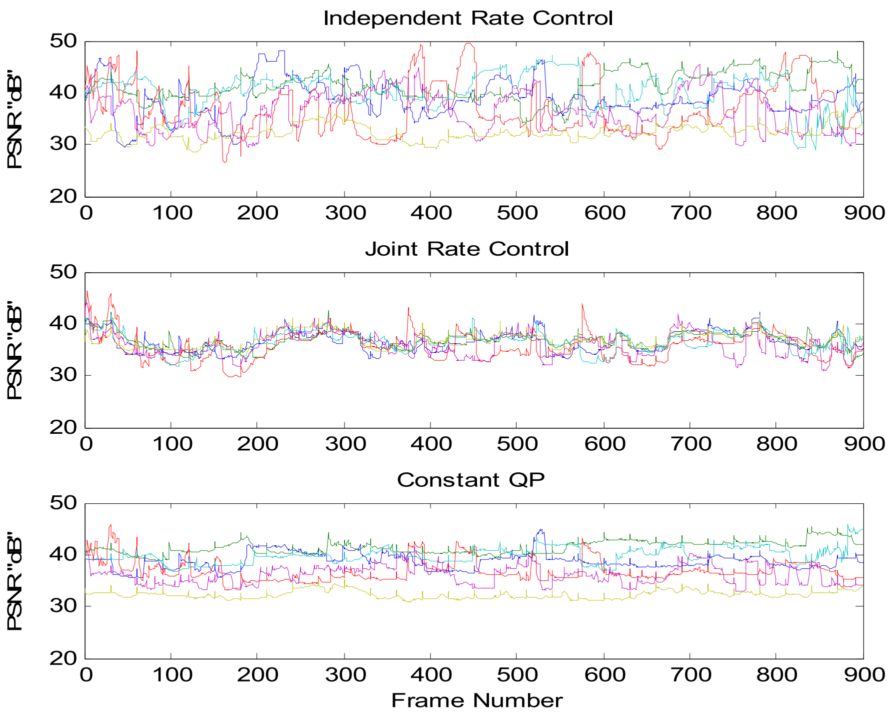

Figure 21: PSNR of encoded bit stream by three encoding methods

4.2. JOINT VIDEO ENCODING WITH INDEPENDENT QUALITY OF SERVICES

A number of existing joint video encoding methods were studied and a new joint encoding method was proposed in [P8]. All these methods try to share the available bandwidth, according to the coding complexities of video sources, and to provide a similar level of quality for the encoded video. In these methods, it is impossible to encode bit streams with different QoS (Quality of Service) levels. However, looking at the joint video encoding as a StatMux method, it should be possible to multiplex bit streams with different QoS levels or bit rates and to utilize the advantages of StatMux. In practice, different QoS levels may be applied according to discrimination of contents or receivers.

A novel joint encoding and StatMux method is proposed in [P9]. In the proposed system the encoded bit streams can have different average long-term bit rates while they can benefit from StatMux in the short-term to decrease end-to-end delay of a broadcast system. The proposed method is implemented using a joint rate control system. Figure 22 shows the simplified block diagram of the rate control system. The rate control system utilizes a fuzzy joint rate controller (JRC) to decrease the buffering delay of the aggregated bit stream and thereafter end-to-end delay. It also utilizes a number of independent rate controllers (IRCi), one for each encoder, to provide a long-term target bit rate for each bit stream. The proposed system is a real-time encoding system without any look ahead. It has a very low degree of computational complexity.

As shown in the block diagram, the encoded bit streams are multiplexed and moved to a virtual joint buffer. Data are removed from the joint buffer with a constants bit rate appropriate to the transmission channel. The occupancy of the joint buffer is used as a feedback signal by the JRC. The JRC controls the variations in the bit rate of the aggregated bit stream to guarantee the broadcasting of bit streams with a limited end-to-end delay through a channel with a defined bandwidth. For a given transmission bandwidth, less variations in the bit rate of the aggregated bit stream means shorter end-to-end delay for broadcast services.

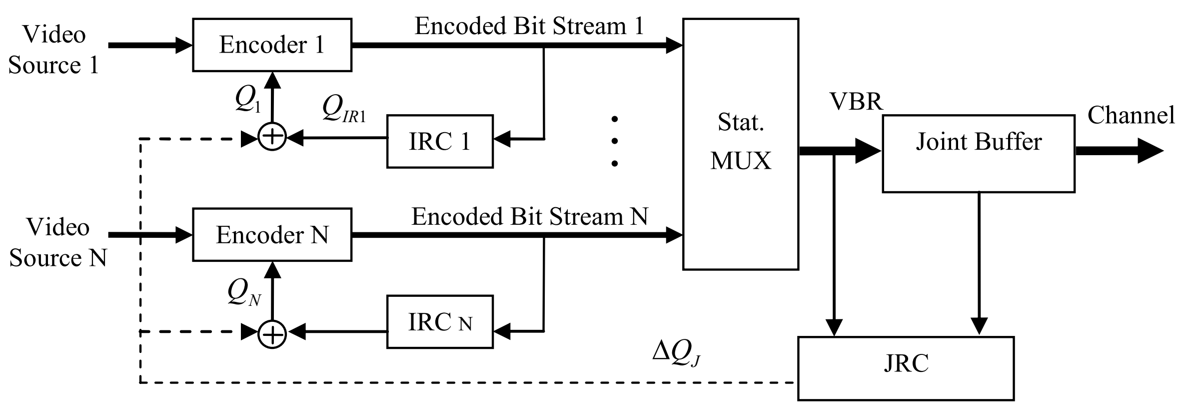

Figure 22: Block diagram of proposed joint encoding and RCA

The video encoders are controlled by the QP per picture basis. The QP is mainly controlled by the IRCs while the JRC may add a small positive or negative value to the QP according to the occupancy of the joint buffer.

$$Q_n = Q_{IRCn} + \Delta Q_J \ , \tag{4.7}$$

where Q_n denotes the used QP by the n^{th} encoder. Q_{IRCn} is the QP calculated by the n^{th} independent rate controller and ΔQ_J stands for the output of JRC. Any VBR rate control algorithm such as the semi-fuzzy RCA can be used as IRC. The output of JRC is computed according to the occupancy of the joint buffer and the bit rate of the aggregated bit stream. More details about the proposed control system are presented in the sequel.

Considering the virtual buffer at the receiver side, the buffer occupancy is updated after encoding a series of corresponding pictures (m^{th} picture of each source) as:

$$O_B(m) = O_B(m-1) - \sum_{i=1}^{N} B_i + R \ , \tag{4.8}$$

where O_B denotes the buffer occupancy. B_i stands for the consumed bits by the encoded picture of the i^{th} source. R represents the channel bandwidth.

Joint Rate Controller

The output of JRC is computed by a fuzzy controller. While each bit stream uses an independent controller with a buffer constraint, without any more control, the aggregated bit stream is constrained by a joint buffer with a size equal to the sum of the sizes of the buffers used by IRCs. The idea is to use a virtual buffer with the size of S_B as small as possible and then use the JRC to operate only when the buffer condition is critical. Therefore, the fuzzy controller has been designed such that the JRC has a non-zero value only when the buffer state is critical. This minimizes the interaction between encoders and as well as the variations in the quality of encoded bit streams. The fuzzy controller has two input signal as:

$$x_1 = O_B / S_B \ , \tag{4.9}$$

$$x_2 = \frac{F \sum_{i=1}^{N} B_i}{\sum_{i=1}^{N} \left(R_i \Big/ \left(1 + \frac{X_{IPi} - 1}{I_{Ii}} \right) \right)} \ , \tag{4.10}$$

where R_i denotes the target bit rate of the i^{th} bit stream. I_{Ii} represents the interval of frequent IDR pictures in the i^{th} bit stream. F stands for the frame rate of bit streams and X_{IPi} indicates the relative coding complexity of IDR pictures to P pictures in the i^{th} bit stream.

All the fuzzy rules are summarized in Table 4. The content of the table specifies the output of the controller. The letters H, L, M and V correspond to linguistic specifications of High, Low, Medium and Very. Seven and three membership functions for the two inputs x_1 and x_2 as shown in Figure 23 are employed. The desired central values for the output of the fuzzy system corresponding to VL, L, ML, M, MH, H and VH in Table 4 are -3, -2, -1, 0, 1, 2, and 3, respectively. A fuzzy system with two inputs, product inference engine, singleton fuzzifier, and centre average defuzzifier as (2.18) is used. Moreover, the output of the fuzzy system is tuned adaptively according to the buffer size as in (2.19).

Simulation Results

The proposed joint video encoding and multiplexing method has been evaluated in comparison with an independent encoding scenario in a DVB-H application. Simulation results show that the proposed joint encoding method can considerably reduce end-to-end delay (about 60%) and buffering requirements (about 40%) at the receivers without any cost in the average quality of encoded video.

TABLE 4. SUMMARIZATION OF THE IF-THEN FUZZY RULES

x_2 \ x_1	**VL**	**L**	**ML**	**M**	**MH**	**H**	**VH**
H	VH	H	MH	M	M	M	ML
M	H	MH	M	M	M	ML	L
L	MH	M	M	M	ML	L	VL

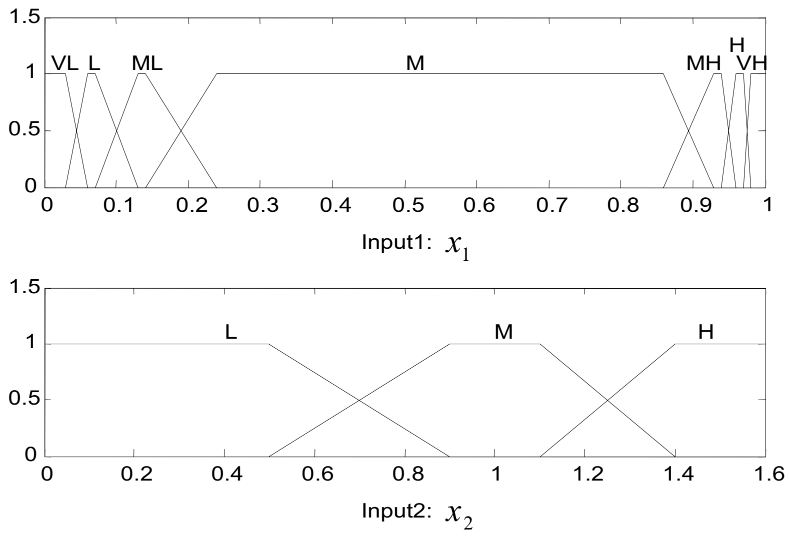

Figure 23: Membership function of the linguistic variables

Chapter 5

Conclusions

Digital video contents are going to consume the main part of the bandwidth in different communication networks. Several video coding standards have been defined to compress the amount of data that is required for presentation of video contents. However, the widely used hybrid video coding standards only define the syntax of the bit stream and the decoding process. The encoding process is not standardized to allow flexible implementations for different applications. A number of video encoding algorithms and tools were introduced in this thesis that are targeted for real-time variable bit rate video applications especially for video streaming applications.

In Chapter 2, first a video rate control algorithm for variable bit rate applications was proposed that can provide a high performance in real-time applications with a low degree of complexity. In this algorithm two reference points corresponding to constant bit rate and constant quality rate control were used to build a reference point for variable bit rate. By this technique the proposed rate control algorithm can be tuned for a wide range of applications from constant bit rate to constant quality applications. Then, a general rate distortion model and a coding complexity measure were introduced that can be used in different video coding standards for the rate control. Moreover, bit allocation algorithms for variable bit rate applications were proposed that can improve the average video quality by reducing the propagation of quantization error. Finally, in a new approach a semi-fuzzy rate control algorithm for variable bit rate applications is introduced that can provide a high performance with a very low degree of computational complexity in a wide range of applications and on different video coding standards.

In Chapter 3, the essential parameters related to video streaming over DVB-H application were studied and a video encoding and splicing method was proposed that can considerably decrease end-to-end delay of a broadcast system.

In chapter 4, novel joint video encoding and statistical multiplexing methods based on fuzzy logic control are introduced that can considerably improve the quality of compressed video and can reduce end-to-end delay of a broadcast system. The reduction of end-to-end delay can be interpreted as a reduction of transmission bandwidth.

As a new approach, fuzzy logic controllers are utilized successfully in conjunction to some other tools and controllers based on the conventional video encoding methods in different video encoding scenarios. The new approach can be used for multi-layer scalable video encoding and also for multi-view video encoding.

Bibliography

ANDREWS H.C., 1968, “Entropy considerations in the frequency domain,” *Proceedings of the IEEE,* Vol. 56, No.1, pp. 113-114, Jan. 1968.

BASHFORTH B.N., WILLIAMSON C.L., 1998, "Statistical multiplexing of self-similar video streams: simulation study and performance results," IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, pp. 119-126, July 1998.

BERGER T., 71, *Rate Distortion Theory*, NJ: Prentice-Hall, Englewood Cliffs, 1971.

BJONTEGAARD G., 2001, “Calculation of average PSNR differences between RD-curves,” *ITU-T Video Coding Experts Group (VCEG)*, Document VCEG-M33, Austin, Texas, USA, 2-4 Apr. 2001.

BOROCZKY L., NGAI A. Y. , WESTERMANN E. F., 2000, “Joint rate control with look-ahead for multi-program video coding,” *IEEE Transactions on Circuits and Systems for Video Technology,* , Vol. 10, No. 7, pp. 1159-1163, Oct. 2000.

BOYCE J. M., TOURAPIS A. M., 2005, “Fast Efficient Channel Change [set-top box applications],” *IEEE Int. Con. on Consumer Electronics (ICCE)*, 8-12 Jan. 2005.

CHANG C.Y., LIN T., CHAN D.Y., HUNG S. H., 2006, “A low complexity rate-distortion source modeling framework,” *Int. Con. Acoustics, Speech, and Signal Processing (ICASSP ‘06)*, pp. 929-932, May, 2006.

CHELLAPPA V., COSMAN P.C., VOELKER G.M., 2004, “Dual frame motion compensation with uneven quality assignment,” *IEEE Data Compression Conference (DCC),* pp. 262-271, 2004.

CHIANG T., ZHANG Y.-Q., 1997, “A new rate control scheme using quadratic rate distortion model*,” IEEE Transaction On Circuits and System for Video Technology*, Vol. 7, No.1, pp. 246-250, Feb.1997.

CHIANG T., ZHANG Y.-Q., LEE J., MARTUCCI S., PETERSON H., SODAGAR I., SURYADEVARA R., Wine C., 1995, “A rate control scheme using a new rate distortion model,” *Dalas, TX: JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG 95/0436*, Nov. 1995.

DAI M., LOGUINOV D., RADHA H. M., 2006, "Rate-Distortion Analysis and Quality Control in Scalable Internet Streaming*," IEEE Transactions on Multimedia*, Vol. 8, No 6, pp. 1135-1146, Dec. 2006.

DAI M., LOGUINOV D., RADHA H., 2003, "Statistical analysis and distortion modeling of MPEG-4 FGS," *IEEE Int. Conf. on Image Processing (ICIP)*, pp. 301-304, Sept. 2003.

ETSI (EUROPEAN TELECOMMUNICATIONS STANDARDS INSTITUTE), 2004 b, Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals (DVB-H), EN302304, V1.1.1, 2004.

ETSI (EUROPEAN TELECOMMUNICATIONS STANDARDS INSTITUTE), 2004 c, Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television, EN300744, V1.5.1, 2004.

ETSI (EUROPEAN TELECOMMUNICATIONS STANDARDS INSTITUTE), 2006 d, Digital Video Broadcasting (DVB); IP Datacast over DVB-H: Content Delivery Protocols, TS102472, V1.1.1, 2006.

ETSI (EUROPEAN TELECOMMUNICATIONS STANDARDS INSTITUTE), 2006 e, Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in DVB services delivered directly over IP protocols, TS102005, V1.2.1, 2006.

FARBER N., GIROD B., 1997, "Robust H.263 Compatible Video Transmission for Mobile Access to Video Servers," *Proc. IEEE Int. Con. on Image Processing (ICIP)*, Santa Barbara, CA, USA, vol. 2, pp. 73-76, Oct. 1997.

FARIA G., HENRIKSSON J. A., STARE E., TALMOLA P., 2006, "DVB-H: Digital Broadcast Services to Handheld Devices," *Proceeding of the IEEE*, V. 94, No. 1, Jan. 2006

GARDOS T. R., 1997, "Video Codec Test Model, Near-Term, Version 8 (TMN8)," *ITU - Telecommunications Standardization Sector*, Document Q15-A-59, Portland, 24-27 June 1997.

GERSHO A, GRAY R. M., 1992, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.

GISH H., PIERCE J. N., 1968, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory,* Vol. 14, No.5, pp. 676–683, Sept. 1968.

HANG H.-M., CHEN J.-J., 1997, "Source model for transform video coder and its application-part I: fundamental theory," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 7, No. 2, pp. 197-211, Apr. 1997.

HE Z., CHEN T., 2003, "Linear rate control for JVT video coding," *International Conference on Information Technology: Research and Education (ITRE)*, pp. 65-68, Aug. 2003.

HE Z., KIM Y., MITRA S. K., 2001a, "A linear source model and a unified rate control algorithm for H.263 / MPEG-2 / MPEG-4," *Proc. of ICASSP*, Salt Lake City, Utah, May 2001.

HE Z., KIM Y., MITRA S. K., 2001b, "A Unified Rate-Distortion Analysis Framework for Transform Coding," *IEEE Transactions on Circuits and Systems for Video Technology,* Vol.11, No. 12, Dec. 2001.

HE Z., KIM Y., MITRA S. K., 2002, "Optimum Bit Allocation and Accurate Rate Control for Video Coding via ρ-Domain Source Modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 12, No. 10, Oct. 2002.

HE Z., MITRA S. K., 2001, "A Unified Rate-Distortion Analysis Framework for Transform Coding*," IEEE Trans. Circuits and Systems for Video Technology*, Vol. 11, Dec. 2001.

ISO/IEC, 1993, ISO/IEC JTC/SC29/WG11/N0400 MPEG93/457, Test Model 5, TM5, Apr. 1993.

ISO/IEC JTC1, 1999, "Coding of Audio-Visual Objects-Part 2: Visual," ISO/IEC 14 496-2 (MPEG-4 visual version 1), 1999.

ISO/IEC JTC1, ITU-T, 1994, "Generic Coding of Moving Pictures and Associated Audio Information - Part 2: Video," ITU-T Recommendation H.262-ISO/IEC 13 818-2 (MPEG-2), 1994.

ISO/IEC JTC1/SC29/WG11 N3908, 2001, "MPEG-4 Video Verification Model version 18.0," Pisa, Jan. 2001.

ITU-T, 1990, "Video Codec for Audiovisual Services at px64 Kbit/s Version 1," ITU-T Recommendation H.261, 1990.

ITU-T, 1995, "Video Coding for Low Bitrate Communication Version 1," ITU-T Recommendation H.263, 1995.

ITU-T, 1997, "Video Codec Test Model, Near-Term, Version 8 (TMN8)", Video Coding Experts Group, 1997.

JAGMOHAN A., RATAKONDA K., 2003, "Mpeg-4 One-Pass VBR Rate Control for Digital Storage," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, pp. 447-452, May 2003.

JAYANT N. S., NOLL P., 1984, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice Hall, 1984.

JVT, 2003, "Advanced video coding for generic audiovisual services," *ITU-T Recommendation International Standard of Joint Video Specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC),* JVT-G050, 2003.

KAMACI N., ALTUNBASAK Y., MERSEREAU R. M., 2005, "Frame bit allocation for the H.264/AVC Video Coder Via Cauchy-Density-Based Rate and Distortion Models," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 15, No. 8, pp. 994-1006, Aug. 2005.

KARCZEWICZ M., KURCEREN R., 2003, "The SP- and SI-frames Design for H.264/AVC", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, Jul. 2003.

KIM Y. K., HE Z., MITRA S. K., 2001, "A Novel Linear Source Model and a Unified Rate Control Algorithm for H.263/MPEG-2/MPEG-4," *IEEE Int. Con. on Acoustics, Speech, and Signal Processing (ICASSP),* Vol.3, pp. 1777-1780, May 2001.

KIM Y., PYUN J.-Y., KIM H.-S., PARK S.-H., KO S.-J., 2003, "Efficient real-time frame layer rate control technique for low bit rate video over WLAN," *IEEE Transactions on Consumer Electronics,* Vol. 49, No. 3, pp. 621-628, Aug. 2003.

KONDO S., FUKUDAA H., 1997, "A Real-time Variable Bit rate MPEG2 Video Coding Method for Digital Storage Media," *IEEE Trans. Consumer Electronics*, Vol. 43, pp. 537-543, Aug. 1997.

KORNFELD M., 2004, "DVB-H - The Emerging Standard for Mobile Data Communication," *IEEE Int. Symposium on Consumer Electronics,* pp. 193-198, Sep. 2004.

LADEBUSCH U., LISS C. A., 2006, "Terrestrial DVB (DVB-T): a broadcast technology for stationary portable and mobile use," *Proceedings of the IEEE*, Vol. 94, No. 1, pp. 183-193, Jan. 2006.

LAKSHMAN T. V., ORTEGA A., REIBMAN A. R., 1998, "VBR video: tradeoffs and potentials," *Proc. IEEE*, Vol. 86, No. 5, pp. 952-973, May 1998.

LEE C.C., 1990a, "Fuzzy logic in control systems: fuzzy logic controller. Part I," *IEEE Transactions on Systems, Man and Cybernetics*, Vol.20, No. 2, pp. 404-418, March-Apr. 1990.

LEE C.C., 1990b, “Fuzzy logic in control systems: fuzzy logic controller. Part II,” *IEEE Transactions on Systems, Man and Cybernetics*, Vol.20, No. 2, pp. 419-435, March-Apr. 1990.

LEE H. J., CHAING T., ZHANG Y. Q., 2000, “Scalable Rate Control for MPEG-4 Video,” *IEEE Trans. Circuits and Systems for Video Technology,* Vol. 10, pp. 878-894, Sep. 2000.

LIN J., ORTEGA A., 1998, “Bit-rate Control using piecewise approximation rate-distortion characteristics,” *IEEE Trans. Circuits and Systems for Video Technolgy*, Vol. 8, No. 4, pp. 446-459, Aug. 1998.

MAGLARIS B., ANASTASSIOU D., SEN P., KARLSSON G., ROBBINS J.D., 1998, "Performance models of statistical multiplexing in packet videocommunications," IEEE Transactions on Communications, Vol. 36, No. 7, pp. 834-844,July 1998.

MA S., GAO W., LU Y., 2005, “Rate-Distortion Analysis for H.264/AVC Video Coding and its Application to Rate Control,” *IEEE Transactions on circuits and Systems for Video Technology,* Vol. 15, No. 12, pp. 1533-1544, Dec. 2005.

MA S., GAO W., WU F., LU Y., 2003a, “Rate Control for Advance Video Coding (AVC) Standard,” *Int. Symposium on Circuits and Systems (ISCAS)*, Vol. 2 , pp. 892-895, May 2003.

MA S., GAO W., WU F., LU Y., 2003b, “Rate control for JVT Video Coding Scheme with HRD Considerations,” *Int. Con. on Image Processing (ICIP),* Vol. 3, pp. 793-6, Sep 2003.

MALVAR H.S., HALLAPURO A., KARCZEWICZ M., KEROFSKY L., 2003, “Low-complexity transform and quantization in H.264/AVC,” *IEEE Transactions on Circuits and Systems for Video Technology,* Vol. 13, No. 7, pp. 598-603, July 2003.

MAY G., 2004, “The IP Datacast System - Overview and Mobility Aspects,” *IEEE Int. Symposium on Consumer Electronics,* pp. 509-514, Sep. 2004.

MOSCHENI F., DUFAUX F., NICOLAS H., 1993, “Entropy criterion for optimal bit allocation between motion and prediction error information,” *Int. Conf. Visual Communications and Image Processingin (VCIP)*, Nov. 1993, pp. 604-612.

NAVAKITKANOK P., ARAMVITH S., 2004, “Improved Rate Control for Advanced Video Coding (AVC) Standard Under Low Delay Constraint,” *International Conference on Information Technology: Coding and Computing (ITCC),* Vol.2, pp. 664-668, 2004.

NGAN K. N., MEIER T., CHEN Z., 2003, “Improved Single-Video-Object Rate Control for MPEG-4,” *IEEE Trans. Circuits and Systems for Video Technology,* vol. 13, pp. 385–393, May 2003

NGUYEN A. G., HWANG J. N., 2002, "SPEM online rate control for real time streaming video," *IEEE Int. Conf. Information Technology* , pp. 65-70, 2002.

PAN F., LI Z., LIM K., FENG G., 2003, "A study of MPEG-4 rate control scheme and its improvements," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No.5, pp. 440-446, May 2003.

PAO I. M., SUN M. T., 1998, "A rate-control scheme for streaming video encoding," *Signals, Systems & Computers, Conference Record of the Thirty-Second Asilomar Conference on* , Vol.2, pp.1616-1620, 1-4 Nov. 1998.

PICKERING M.R., ARNOLD J. F., 1994, "A Perceptually Efficient VBR Rate Control Algorithm," *IEEE Transaction on Image Processing*, Vol. 3, pp. 527-532, Sept. 1994.

POSTEL J., 1981, "RFC 791: Internet protocol*," The Internet Engineering Task Force (IETF)*, Sept. 1981.

REED E. C., DUFAUX F., 2001, "Constrained Bit-Rate Control for Very Low Bit-Rate Streaming-Video Applications," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 11, pp. 882-889, July 2001.

REED I. S., SOLOMON G., 1960, "Polynomial Codes over Certain Finite Fields," *SIAM Journal of Applied Mathematics*, Vol. 8, pp. 300-304, 1960.

RIBAS-CORBERA J. , NEUHOFF D. L., 1998, "Optimizing block size in motion compensated video coding," J. Electron. Imaging, Vol. 7, pp. 155-165, Jan. 1998.

RIBAS-CORBERA J. , NEUHOFF D., 1996, "On the optimal motion vector accuracy for block-based motion-compensated video coders," in Proc. IS&T/SPIE Dig. Video Comp.: Alg. & Tech., San Jose, CA, pp. 302-314, Jan.-Feb. 1996.

RIBAS-CORBERA J., CHOU P.A., REGUNATHAN S.L., 2003, "A generalized hypothetical reference decoder for H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology,* Vol. 13, No. 7,pp. 674-687, July 2003.

RIBAS-CORBERA J., LEI S., 1999, "Rate control in DCT video coding for low-delay communications," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, Issue 1,pp. 172-185, Feb. 1999.

RICHARDSON I. E. G., 2005, "*H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia*," John Wiley and Sons, 2005.

RONDA J.I., ECKERT M., JAUREGUIZAR F., GARCIA N., 1999, "Rate control and bit allocation for MPEG-4," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 8, pp. 1243-1258, Dec. 1999.

SETTON E., GIROD B., 2005, “Video streaming with SP and SI frames”, *Proceeding of Visual Communications and Image Processing (SPIE-VCIP)*, Beijing, China, July 2005.

SHANNON C. E., 1948, “A Mathematical Theory of Communication,” *Bell System Technical Journal,* Vol. 27, pp. 379–423, 623–656, July, Oct., 1948.

SMOOT S.R., ROWE L.A., 1996a, “Study of DCT Coefficient Distributions,” *SPIE Symposium on Electronic Imaging*, Feb. 1996.

SMOOT S.R., ROWE L.A., 1996b, “Laplacian Model for AC DCT Terms in Image and Video Coding,” *Proc. Ninth Image and Multidimensional Signal Processing workshop,* Mar. 1996.

SONG B. C., CHUN K. W., 2003, “A one-Pass Variable Bit-Rate Video Coding for Storage Media,” *IEEE Trans. Consumer Electronics*, Vol. 49, pp. 689-692, Aug. 2003.

SONG H., KUO C.-C.J., 1999, “Rate control of a region-based H.263 video codec under time-varying channels,” *IEEE 3rd Workshop on Multimedia Signal Processing,* pp. 339-344, 13-15 Sep. 1999.

SONG H., KUO C.-C.J., 2001, “Rate control for low-bit-rate video via variable-encoding frame rates,” *IEEE Transactions on Circuits and Systems for Video Technology,* Vol.11 , No. 4, pp. 512–521, Apr. 2001.

SULLIVAN G. J., WIEGAND T., 1998, “Rate-distortion optimization for video compression,” *IEEE Signal Processing Magazine*, Vol.15, Issue 6, pp. 74-90, Nov. 1998.

SULLIVAN G., WIEGAND T., LIM K.P., 2003, “Joint Model Reference Encoding Methods and Decoding Concealment Methods,” *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG*, Document JVT-I049, San Diego, USA, Sep. 2003.

SULLIVAN G.J., WIEGAND T., 2005, “Video Compression - from concepts to the H.264/AVC standard,” *Proceeding of The IEEE*, Vol. 93, No. 1, pp. 18-31, Jan. 2005.

TAKAMURA S., KOBAYASHI N., 2002, “MPEG-2 one-pass variable bit rate control algorithm and its LSI implementation,” *IEEE Int. Conf. Image Processing*, pp. 942-945, 2002.

TSAI J.-C., 2005, “Rate control for low-delay video using a dynamic rate table,” *IEEE Transactions on Consumer Electronics*, Vol. 15, No. 1, pp. 133- 137, Jan. 2005.

TSAI J.-C., Hsieh C.-H., 2004, “Modified TMN8 Rate Control for Low-Delay Video Communications,” *IEEE Transaction On Circuits and System for Video Technology*, Vol. 14, No. 6, pp. 864-868, June 2004.

TSANG D.H.K., BENSAOU B., LAM S.T.C., 1998, “Fuzzy-based rate control for real-time MPEG video,” *IEEE Transactions on Fuzzy Systems*, Volume 6, Issue 4, pp. 504-516, Nov. 1998

VARSA V., KARCZEWICZ M., 2001, "Long Window Rate Control for Video Streaming," *Proceedings of the 11th International Packet Video Workshop*, Kyungju, Korea, pp. 154- 159, Apr. 2001.

VETRO A., SUN H., WANG Y., 1999, "MPEG-4 rate control for multiple video objects," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No.1, pp. 186-199, Feb. 1999.

WANG L. X., 1994, "*Adaptive Fuzzy System and Control: Design and Stability Analysis*," NJ: Prentice-Hall, Englewood Cliffs, 1994.

WANG L., VINCENT A., 1999, "Bit allocation and Constraints for Joint Coding of Multiple Video Programs," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 6, pp. 949-959, Sep. 1999.

WANG L.X., 1993, "Stable adaptive fuzzy control of nonlinear systems," *Fuzzy Systems, IEEE Transactions on*, Vol. 1, No. 2, pp. 146-155, May 1993.

WIEGAND T., GIROD B., 2001, "Lagrange Multiplier Selection in Hybrid Video Coder Control," *IEEE International Conference on Image Processing*, Vol. 3, pp. 542-545, Oct. 2001.

WIEGAND T., SCHWARZ H., JOCH A., KOSSENTINI F., SULLIVAN G.J., 2003a, "Rate-Constrained Coder Control and Comparison of Video Coding Standards," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp. 688-703, July 2003.

WIEGAND T., SULLIVAN G.J., BJNTEGAARD G., LUTHRA A., 2003b, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits and System for Video Technology*, pp. 560-576, July 2003.

XIONG H., SUN J., YU S., LUO C., ZHOU J., 2004, "Design and Implementation of Multiplexing Rate Control in Broadband Access Network TV Transmis-sion System," *IEEE Transactions on Consumer Electronics*, Vol. 50, No. 3, pp. 849-855, Aug. 2004.

YANG X., LIN W., LU Z., LIN X., RAHARDJA S., ONG E.P., YAO S., 2005, "Rate control for videophone using local perceptual cues," *IEEE Transactions on Circuits and Systems for Video Technology,* Vol. 15, No. 4, pp. 496-507, Apr. 2005.

YONEYAMA A., YANAGIHARA H., NAKAJIMA Y., 2001, "One-pass VBR MPEG encoder using scene adaptive dynamic GOP structure," *IEEE International Conference on Consumer Electronics (ICCE),* pp. 174-175, June 2001.

ZADEH L. A., 1965, "Fuzzy Sets," *Information and Control*, 8(3), pp. 338-353, 1965.

ZHANG ZHI-LI, KUROSE J., SALEHI J.D., TOWSLEY D., 1997, "Smoothing, statistical multiplexing, and call admission control for stored video," IEEE Journal on Selected Areas in Communications, Vol. 15, No. 6, pp. 1148-1166, Aug. 1997.

ZHANG X. M., VETRO A., SHI Y. Q., SUN H., 2003, "Constant Quality Constrained rate Allocation for FGS-Coded Video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, pp. 121-130, Feb. 2003.

ZHAO L., KIM J. W., KUO C. C., 2002, "MPEG-4 FGS video streaming with constant-quality rate control and differentiated forwarding," *SPIE Visual Communications and Image Processing (VCIP)*, pp. 230-241, Jan. 2002.

Publications